UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA DE LISBOA

Flexibility and Control of Attention

Inês Antunes de Santa Ana

Curso de Mestrado em Neurociências

Lisboa, 2015

UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA DE LISBOA

Flexibility and Control of Attention

Inês Antunes de Santa Ana

Dissertação orientada pelo Prof. Doutor Tiago Vaz Maia

A impressão desta dissertação foi aprovada pelo Conselho Científico da Faculdade de Medicina de Lisboa em reunião de 26 de Maio de 2015.

Whether you think you can or you think you can't
either way you are right!

Henry Ford

ACKNOWLEDGMENTS

I would like to express my special gratitude to Professor Tiago Vaz Maia, my supervisor, for guiding me through this journey and helping me grow as a scientist.

A warmly thank you to ALL my colleagues that have always been there to support me when I needed and encouraged me to work harder. It is a true pleasure to work with you!

A special thanks goes to my family who always supports my decisions.

And a huge thank you to Pedro who keeps me motivated and makes me see the beauty of life, probably because he makes my life much more beautiful.

ABSTRACT

Attention is an important cognitive process that enables us to concentrate on relevant stimuli while ignoring irrelevant or distracting ones. It mediates our ability to concentrate on a task or to focus on the most important information. This thesis is a conjunction of two experiments that assessed specific aspects of attention.

The first experiment aimed to investigate the occurrence of ironic effects when subjects are voluntarily suppressing their attention to a set of distracting images. This experiment is based on the theory of ironic effects of mental control. According to this theory, trying to control our mind under cognitive load can result in the exact opposite outcome, what is called the ironic effect. A novel cognitive task was designed to evaluate the occurrence of ironic effects on active suppression of distracting images presented on the background of the screen while the subjects performed a Go/NoGo task. Fifty-six adult subjects completed this task in two different environments: half of the subjects were tested in quiet conditions, and the other half performed the test in a room with moderate noise. The results show opposite effects of attention suppression for the two testing groups: while the subjects tested in quiet conditions benefited from actively suppressing their attention to the distracting images, the group tested under moderate noise conditions showed a worse performance when they actively avoided paying attention to the same images, suggesting the occurrence of ironic effects on the latter group.

The second experiment aimed to develop a new neurocognitive task to evaluate attention shifting. Attention shifting is an important process of mental flexibility that mediates our ability to shift attention from previously relevant stimuli to new relevant ones. It has been assessed by cognitive tasks that also include other important cognitive processes, such as feedback processing and reinforcement learning. The newly developed task evaluates attention shifting without the confounders present in these previous tasks: it does not provide feedback to the participants, but rather relies on uncertainty to guide the participants' responses. Forty-five adults completed this task. The results show that the majority of the participants could detect the changes in context and effectively shift their attention to the new relevant stimulus. Therefore, our task is a good measure of attention shifting and will be applied in future studies to assess attention shifting impairments in neuropsychiatric disorders as attention-deficit/hyperactivity disorder and obsessive-compulsive disorder.

Key terms

Attention; ironic effects; attention suppression; attention shifting; mental flexibility.

RESUMO

A atenção é um importante processo cognitivo que nos permite concentrar em estímulos relevantes e ignorar estímulos irrelevantes ou distratores. A atenção medeia a nossa capacidade de concentração numa tarefa ou de focar a informação mais importante em cada contexto. A presente dissertação é o resultado de dois estudos que pretendem investigar diferentes aspetos da atenção.

O primeiro estudo tinha como objetivo investigar a ocorrência de efeitos irónicos na supressão voluntária da atenção para um conjunto de imagens distratoras. Esta experiência baseia-se na teoria de efeitos irónicos do controlo mental, a qual defende que tentar controlar a mente sob esforço pode ter um desfecho exatamente oposto ao desejado, o que é designado por efeito irónico. Uma nova tarefa cognitiva foi desenhada para avaliar a ocorrência de efeitos irónicos durante a supressão ativa de imagens distratoras apresentadas no fundo do ecrã enquanto os participantes realizavam uma tarefa de Go/NoGo. Cinquenta e seis adultos participaram neste estudo. Metade dos participantes realizou a tarefa num ambiente silencioso, enquanto a outra metade realizou a tarefa em ambientes com algum ruído. Os resultados mostram efeitos opostos para os dois grupos de participantes: para os participantes testados em ambientes sem ruído, a supressão das imagens distratoras beneficiou o seu desempenho na tarefa, enquanto que para os participantes testados sob ruído, a supressão das imagens afetou negativamente o seu desempenho, sugerindo a ocorrência de efeitos irónicos neste último grupo de participantes.

O segundo estudo realizado pretendia desenvolver uma tarefa cognitiva inovadora para avaliar a atenção alternada. A atenção alternada é um processo essencial da flexibilidade mental e medeia a nossa capacidade de alternar a atenção entre estímulos anteriormente relevantes para novos estímulos relevantes. Esta capacidade tem sido testada com tarefas que envolvem também outros processos cognitivos, como o processamento de feedback e a aprendizagem por reforços. Uma nova tarefa foi desenvolvida para investigar atenção alternada sob incerteza. Esta tarefa evita os fatores de confundimento presentes em tarefas semelhantes, uma vez que não dá feedback aos participantes, mas antes usa a incerteza para guiar o seu desempenho. Quarenta e cinco participantes adultos realizaram esta tarefa e a análise completa dos seus resultados foi efectuada. Os dados do estudo mostram que a maioria dos participantes conseguiu detetar as mudanças no contexto da tarefa e alternar a sua atenção para o novo estímulo relevante. Esta tarefa mostrou-se uma boa medida de atenção alternada e será utilizada em estudos futuros no nosso laboratório para explorar défices de atenção alternada em doenças neuropsiquiátricas, como a perturbação de hiperatividade e défice de atenção e a perturbação obsessivo-compulsiva.

Expressões chave

Atenção; efeitos irónicos; supressão de atenção; atenção alternada; flexibilidade mental.

TABLE OF CONTENTS

General introduction and motivation 1

Part 1 3

Introduction 4

Theory of ironic processes of mental control 4

Ironic processes of thought suppression 6

Ironic rebound effects 6

Studies of ironic effects outside thought suppression 7

Functional neuroanatomy of mental control 8

Ironic processes in psychopathology 8

Methods 11

Participants 11

Procedure 11

Experimental task 11

The Go/NoGo paradigm 12

Task division 13

Instructions 14

Questionnaire 15

Data analysis 16

Stimulus discriminability and response bias 17

Questionnaires' responses analysis 20

Results and discussion 21

Demographic data 21

Performance on the ironic effects of attention suppression task 21

Mean reaction times 23

Incorrect responses 25

Response discriminability and bias 32

Ironic rebound effects analysis 35

Questionnaire's responses 38

Final discussion and conclusion 42

Future work 44

Part 2 45

Introduction 46

Mental Flexibility 46

Attention shifting 48

The Posner task 48

The linear maze navigation task: 49

Neurobiology of attention shifting and uncertainty 49

Objective 50

Methods 51

Participants 51

Procedure 51
Experimental task 51
Instructions 53
Data analysis 54
Validity effect 54
Mixed-effects models 54
Parameter estimation 56
Model comparison 57
Results and discussion 58
Demographic data 58
Performance on the attention shifting and uncertainty task 58
Incorrect responses 58
Validity effects 59
Model fitting analysis – mixed-effects models 63
Assessing the fit of the model 69
Final discussion and conclusion 70
Future work 72
References 73
Appendices 83
A. 1.1 77
A. 1.2 78
A. 1.3 79
A. 1.4 80
A. 1.5 81
A. 2.1 82
A. 2.2 83

LIST OF FIGURES

PART 1

Figure 1.1 Task stimuli and fixation point 14
Figure 1.2 Timings of two illustrative trials on the task 15
Figure 1.3 Logit curve 19
Figure 1.4 Signal detection theory. Noise (N) and signal (S+N) distributions 21
Figure 1.5 Go mean reaction times 25
Figure 1.6 NoGo mean reaction times 26
Figure 1.7 Proportion of omission errors 27
Figure 1.8 Proportion of commission errors 27
Figure 1.9 Quantilie-Quantile plot to assess multivariate normality of the data 29
Figure 1.10 Hypothesis-error plots for the three-way interaction condition x block x noise 30
Figure 1.11 Mean within-subjects differences for Go reaction times 31
Figure 1.12 Mean within-subjects differences for NoGo reaction times 32
Figure 1.13 Mean within-subjects differences for omission errors 32
Figure 1.14 Mean within-subjects differences for commission errors 33
Figure 1.15 Canonical discriminant analysis results for the factor condition 34
Figure 1.16 Measures of stimulus discriminability 35
Figure 1.17 Measures of response bias 36
Figure 1.18 Ironic rebound effects for Go mean reaction time 38
Figure 1.19 Ironic rebound effects for NoGo mean reaction time 38
Figure 1.20 Ironic rebound effects for omission errors 39
Figure 1.21 Ironic rebound effects for commission errors 39
Figure 1.22 Subjects mean responses in the questionnaire about the task performance 41
Figure 1.23 Correlogram depicting the correlation between the responses to the questionnaire relative to the allow condition and the subject's performance on the allow condition 42
Figure 1.24 Correlogram depicting the correlation between the responses to the questionnaire relative to the avoid condition and the subject's performance on the avoid condition 43

PART 2

Figure 2.1 Example of a trial on a computational version of the Wisconsin card sorting task 50
Figure 2.2 Timing of an illustrative trial on the task 53
Figure 2.3 Different trial types on the task 53
Figure 2.4 Boxplot for proportions of omission and commission errors in the task 62
Figure 2.5 Boxplots for proportion of omission (a) and commission (b) errors divided by trial type 63
Figure 2.6 Mean validity effects for each sub-block (group of 10 trials) of the task 65
Figure 2.7 Mean validity effects for each block of the task 65
Figure 2.8 Mean validity effects for each sub-block of the task 66
Figure 2.9 Linear model fits for the models with parameters (a) type and trial and (b) type, trial, and shift 70

Figure 2.10 Logarithmic model fits for the models with parameters (a) type and trial and (b) type, trial, and shift ____ 71
Figure 2.11 Logarithmic model fit for the model with parameters type, trial, and validity, with random intercept and slope ____ 72
Figure 2.12 95% prediction intervals on the random effects versus quantiles of the normal distribution ____ 73

LIST OF TABLES

PART 1

Table 1.1 Ironic effects of attention suppression design ____ 14

Table 1.2 Questionnaire about the task performance ____ 15

Table 1.3 Signal detection theory. Possible responses to the stimuli ____ 18

Table 1.4 Demographic characteristics of the participants enrolled in the experiment of ironic effects of attention suppression ____ 21

PART 2

Table 2.1 Attention shifting and uncertainty task design ____ 53

Table 2.2 Model comparison - mixed-effects models ____ 64

LIST OF ABBREVIATIONS

BRIEF – Behavioural Rating Inventory of Executive Functions

fMRI – Functional Magnetic Resonance Imaging

VE – Validity Effect

WBSI – White Bear Suppression Inventory

WCST – Wisconsin card sorting task

GENERAL INTRODUCTION AND MOTIVATION

Attention is a generally intuitive concept, but it is not always easily explained. It can be defined as the cognitive process that allows our brain to focus on particularly relevant stimuli, while ignoring irrelevant contents[1]. Attention is context-dependent and is indispensable to motivate behaviour; it is modulated by both top-down voluntary processes and bottom-up stimulus-driven mechanisms, which influence the way information is processed in our brain.

The ability to appropriately control and effectively use our attention is essential to achieve our goals and met the requirements of different situations. Attentional control is one of the most important achievements of the executive system and is involved in focused attention and attention shifting, two processes continuously required in everyday life. Efficient attentional control is compromised in several psychiatric disorders and further studies on different aspects of attention are important to understand the psychological and neurobiological processes behind it.

The present work aimed to conduct two experiments on distinct properties of attention: the first part of the thesis explores attentional control and is dedicated to the study of ironic effects of attention suppression, while the second part of this thesis investigates attention shifting under uncertainty.

1

IRONIC EFFECTS OF ATTENTION SUPPRESSION

INTRODUCTION

The harder you try, the worse it gets.

Popular saying

When we want to achieve a certain goal, we have to work for it. More importantly, we need to behave according to that goal, which may require self-discipline and self-control. If we want to follow a certain behaviour, it is natural that we try to control ourselves in order to achieve it. And controlling ourselves is a synonym of controlling our mind, by promoting mental states congruent with the desired behaviour and avoiding mental states conflicting with our goals[2]. For instance, it can seem obvious that a person trying to follow a diet might begin by stopping thinking about food, someone trying to abstain from alcohol should start by avoiding alcohol-related thoughts, a person who feels overanxious might mentally try to relax, or a depressed person should avoid sad thoughts and try to think of happy ones.

We believe that we can control ourselves and our own thoughts, and it is a comfortable conviction that suffices everyone from time to time. But we know that controlling our mind, and thus ourselves, is not an easy endeavour, and sometimes the effort we make to control our minds can backfire, leading us in the exact opposite direction from what we were trying to achieve. The study of ironic effects of attention suppression presented in this thesis investigates whether trying to suppress our attention from distracting stimuli does indeed help us to concentrate or only makes things worse.

THEORY OF IRONIC PROCESSES OF MENTAL CONTROL

The theory of ironic processes of mental control was developed by Daniel Wegner in the late 80's[3,4] and tries to explain why mental control frequently results in the exact opposite effect we were trying to achieve. Wegner and his colleagues started to study this phenomenon by asking people to avoid thinking about a white bear. They observed that people would think about the white bear repeatedly as a result of such instructions. In that case, trying to avoid thinking about the white bear is just what makes them think about it. The theory is focused on thought suppression and, indeed, the majority of Wegner's work on ironic processes of mental control addresses the control of thoughts, but this idea can be generalized to any attempt to control our minds.

Wegner's theory recognises that mental control can be tricky and result in ironic processes, suggesting that this effect is due to the architecture of our mental control system. According to this theory, any attempt to control our mind is performed by two processes: (1) an intentional and conscious operating process, which performs an effortful search for mental contents to achieve the desired state of mind, and (2) an unintentional and unconscious ironic monitoring process that automatically searches for mental contents or mental states that signal a failure to produce the desired state of mind. For example, when someone is trying to suppress a thought, the operating process can look for distractors to keep our thoughts away from the forbidden one, while the monitoring process looks for occurrences of the forbidden thought itself[5]. If someone is trying to fall asleep, the operating process might search for signs of fatigue while the monitoring process looks for signs of wakefulness.

The mechanism of mental control works pretty well in most situations, with the two processes functioning in synergy to achieve mental control. That is why we do not always say whatever comes to our mind and have the capacity to ignore the chocolate bar that repeatedly calls for our attention. Nonetheless, there are times when this control fails and we find ourselves doing the exact thing we were trying to avoid. Usually, the monitoring process works silently while the operating process performs its work keeping our mind in the desired state, but these processes can fail, leading to an excessive function of the monitoring process whose effects on mind can supersede those of the operator. For instance, a person who is trying to quit smoking adopts a strategy of suppressing smoking-related thoughts. The operating process will consciously work to produce smoking-free thoughts while the monitoring process unconsciously screens for thoughts of cigarettes, by scanning memories and environmental cues. Whenever the monitoring process finds smoking-related thoughts, it brings them to consciousness in order to restart the operating process, which will again find some distractors to keep the person from thinking about smoking. This strategy might succeed, but it might also happen that suppressing smoke-related thoughts leads to an increase in their occurrence. The continuous search of the monitoring process for smoking-related thoughts can have the undesired effect of increasing the accessibility of the unwanted thoughts. In this case, cigarettes, smoke, tobacco and ashtrays are highlighted by the monitoring process, which becomes counterproductive in avoiding those thoughts. Nonetheless, this is not a problem and will not affect the person's primary goal as long as operating process is working effectively. However, in some circumstances the monitoring process might overtake the operating process, producing not only a failure in mental control but additionally resulting in the exact opposite effect and bringing to mind the unwanted contents.

This is called the ironic effect of mental control and happens more often when people are under cognitive load like stress, time urgency, or in an environment with several distractions. A failure in mental control can happen because the operating process requires effort and available cognitive resources, but those resources are not always available. For instance, if we are under high cognitive load, the operating process can be compromised and ironic processes can arise. When these ironic effects occur, trying to relax can get us more anxious, trying to fall asleep can keep us awake for longer periods, and trying not to think about food can make us obsessed with our next meal. Engaging in an innocent program of mental control trying to improve ourselves might be the fastest way to fail and worsen the problem we wanted to overcome.

Mental control can be easily related to thought control. So we will start by addressing ironic processes of thought suppression – trying not to think about something. And trying not to think about something can be the most effective way of non-stop thinking about it! From now on I would like to ask you not to think about the **pink panther**.

Source: http://www.pinkpanther-wine.com/images/Pink-Panther_top.png

IRONIC PROCESSES OF THOUGHT SUPPRESSION

Evidence suggests that intentional suppression of thoughts ironically enhances the accessibility of the unwanted thought[5,6], specifically when people are under mental load, and this prediction has been corroborated repeatedly. Supressing a specific thought under mental load increases the accessibility of that thought[7], intended concentration under load enhances the accessibility of distractors[8], controlling our mood under load leads to the occurrence of thoughts related to the unwanted mood and actual experience of the unwanted mood[6], trying to relax under load produces arousal[9], trying to sleep under load produces wakefulness[10], effortful forgetting under load yields remembering[11], trying not to perform specific movements under load will make us do the precise movement we were trying to avoid[12], and supressing pain sensations ironically impairs pain recovery[13].

There are numerous studies that support this theory and evoking all of them would be a very long list. However, this effect is not always clear. Instructions to suppress a thought actually reduce self-reported thinking as compared to instructions to concentrate on the thought, or instructions to monitor the thought[7]. Mental control can fail, but most of the times it is indeed successful. Comprehensive reviews on thought suppression found inconsistent ironic effects across different studies. Whereas some studies find an ironic effect while subjects were actively suppressing a thought[14,15], other studies only found ironic effects after the suppression phase[16–18], and there are also studies that do not find any ironic effect of thought suppression[19–21].

IRONIC REBOUND EFFECTS

Pivotal research on the ironic effects of mental control has also shown that efforts to suppress a thought can lead to a later thought rebound, what was called the ironic rebound effect. Participants instructed to suppress and then actively express thoughts about a white bear and participants instructed to reversely express and then suppress those thoughts showed distinct results:

participants who firstly suppressed the thought reported more thought occurrences on the expression phase than participants who expressed first and suppressed secondly[22].

A meta-analysis conducted by Abramowitz, Tolin, and Street in 2001[23] of studies of thought suppression found a small to moderate rebound effect of thought suppression but no immediate ironic effect on the frequency of the suppressed thought. Moreover, evidence suggests that ironic rebound effects are stronger for the participants that could successfully suppress the target thoughts during the suppression conditions[4].

STUDIES OF IRONIC EFFECTS OUTSIDE THOUGHT SUPPRESSION

The ironic process theory can be applied to different domains that imply mental control. If we consider that the ironic process is part of the machinery for mental control, it can be found in different situations where we try to control our mind. Since our study is not about thought suppression, I will briefly describe some other studies that addressed ironic effects of mental control in other cognitive processes.

Ironic mood effects

Studies where people try to make themselves happy found that they actually became sad, whereas those people trying to make themselves sad actually experience a happier mood[6].

Ironic effects also arise in the mental control of anxiety. People trying to relax under cognitive load (instructed to rehearse a long number) showed an increase in anxiousness, measured by an increase in skin conductance level, relatively to those people who were performing the same cognitive task but were not instructed to relax[9].

Ironic effects in the control of sleep

Other studies found that people encouraged to "fall asleep as quickly as you can" as they listen to raucous distracting music stay awake much longer than those subjected to the same music but not instructed to quickly fall asleep[10].

Ironic effects in the control of movement

Controlling movement can also be susceptible to ironic effects. A study evaluated subjects' performance while they tried to keep a handheld pendulum from moving in a specific direction. In fact, those subjects made more movements with their pendulum in that direction than the subjects who were simply instructed to hold the pendulum steady, without mentioning any specific direction[12]. In other experiments, subjects tried to avoid overshooting a golf putt and, again, made it more often than the control group of subjects[12].

These studies resemble situations where we are specifically told to be careful. Actions that we perform on a daily basis can fail when someone asks us not to make any mistake. Maybe this is because we feel some pressure and get anxious about doing the action, but there might be ironic effects behind that fact.

Ironic effects in the control of stereotyping

Another study has found ironic rebound effects in the mental control of stereotyping and prejudice. When subjects were asked not to stereotype a skinhead, they were able to avoid stereotypical ideas on their description, and actually used fewer stereotypes than the control people not asked to avoid them. However, the suppression group ended up by showing greater stereotyping after the suppression phase, when they had to sit in the same room with a skinhead[24]. Additionally, people that were trying to forget the stereotypical characteristics of another person were more likely to recall those exact characteristics when under mental load[11].

FUNCTIONAL NEUROANATOMY OF MENTAL CONTROL

Recently, imaging studies have suggested two different brain mechanisms for sustained mental control. The prefrontal cortex is continuously working during the maintenance of a specific cognitive task, whereas the anterior cingulate cortex gets active if additional control is needed due to conflict signals[25–28]. These findings are in line with Wegner's theory of mental control, which predicts that mental control is performed by a sustained process that maintains an active representation of the to-be-avoided concept in mind (the monitoring process), and a transiently active process that is engaged following failures in mental control and provides additional control, helping to successfully return to the suppressing state (the operating process).

In 2007, Wegner and his colleagues[29] conducted an imaging study, using functional magnetic resonance imaging (fMRI) to investigate cortical activity during thought suppression. The subjects were forbidden to think about white bears. During the suppression condition, they were instructed to suppress white bear thoughts and report each time they thought about white bears by pressing a button. In free-thought periods, subjects were allowed to think freely, including white bear thoughts, but were also asked to report every white bear thought by pressing the same button. Results were consistent with previous findings of ironic effects, as the subjects have reported the same amount of white bear thoughts on the suppression and the free-thought conditions. Additionally, fMRI data was consistent with both cognitive and neural models of mental control, with active thought suppression producing both sustained and transient processes that were associated with activation in dorsolateral prefrontal cortex and the anterior cingulate cortex, respectively. Their results provided important insights on the neuroanatomy of thought suppression and have strengthened the initial theory proposed by Wegner.

IRONIC PROCESSES IN PSYCHOPATHOLOGY

Psychopathologies like anxiety, depression, trauma and post-traumatic stress, obsessive-compulsive disorder, sleep disorders, eating disorders, or specific phobia have been linked to ironic processes. Thought suppression has been implicated as an aetiological and/or sustaining factor in these psychiatric disorders[30], because patients often try to overcome their symptoms by suppressing their minds from thinking about them or by trying to inhibit behaviours related to those symptoms.

Even in healthy subjects, the attempt to control the mind can result in laboratory analogues of unwanted mental states that are characteristic of the psychiatric disorders mentioned above. In fact, patients with obsessive-compulsive disorder report more thought occurrences when asked to suppress a neutral thought than anxious or non-anxious control individuals[31], which indicates an enhanced tendency for ironic effects in obsessive-compulsive disorder. It remains to understand whether the higher accessibility to forbidden thoughts in obsessive-compulsive patients is due to their disorder, or if their natural propensity to recall forbidden thoughts lead them to develop the disorder. Either way, we know that thought suppression and mental control are important processes in psychiatric disorders and deserve to be comprehensively studied. Other studies have found higher accessibility of anxiety-related thoughts in anxiety disorders[32], of depression-related thoughts in depression [33] and food- and body-related thoughts in eating disorders[34]. It is reasonable to hypothesise that, by trying to mentally control their problems, these patients' enhanced monitoring process ironically brings their concerns intrusively back to their minds[35].

Another line of evidence comes from experiments that study what happens when mental control is abolished. James Pennebaker and colleagues found that when people are encouraged to express their deepest thoughts and feelings through writing, their psychological and physical health improves[36]. By encouraging people to rescind their constant mental control, the key requirement for the creation of ironic effects is eliminated. After all, disclosing one's personal thoughts to other, or even to oneself, may be the fist step to abandon the demanding quest to control one's own thoughts and emotions. Ironic effects have been hypothesised to explain the recurrence of unwanted thoughts involved in the maintenance of mood and anxiety disorders, since the effort to suppress thoughts related with anxiety or depression can paradoxically enhance the alertness to those thoughts and, consequently, their availability, as well as enduring the emotional reactivity to the thought itself[30].

The idea that trying to control our mind by suppressing specific thoughts can be related with psychopathology is not new and has even been proposed during the twentieth century by Sigmund Freud, who argued that repression played an important role in the development of psychopathology[37]. Also Anna Freud defended this idea, claiming that "Repression is the most dangerous defence mechanism (...) Repression is the basis for the formation of neurosis"[38]. Nonetheless, not everyone shares this point of view on repression. William James[39], for instance, believed that suppressing undesired cognition by shifting to another thought is an important method to regulate consciousness and deal with psychopathology. In fact, Rassin[40] experiments found thought suppression to be an important strategy, especially to control unwanted thoughts. He asked participants to study a very discomforting sentence: "I hope that [a loved one] will soon be in a car accident". Participants were told to either suppress that thought or not to suppress it. In the end, participants who did suppress the thought had fewer thought occurrences (even after the active suppression phase) and felt less distress and less guilty over the thought. This helps to understand the importance of suppression of unwanted thoughts, specifically in disorders like obsessive-compulsive disorder, where suppression represents a useful strategy, like a neutralizing or compulsive act. Patients with obsessive-compulsive disorder tend to use suppression before the thought becomes stronger and they have to perform the compulsion. On the other hand, thought suppression is very important for them, since allowing their obsessional thoughts to occur makes them feel guilty about them and increases their perceived probability of the event coming true. Suppression helps these patients in managing their negative thoughts and the distress they feel about them.

METHODS

PARTICIPANTS

Fifty-six adult subjects participated in this experiment. The majority of the participants were recruited from Universidade de Lisboa, either by direct approach, by randomly asking subjects on libraries and study rooms if they would like to participate, or by email invitation.

All participants were native speakers of Portuguese and performed the task voluntarily after providing written informed consent.

PROCEDURE

The experiment was performed using a laptop on which the subjects performed a cognitive task and answered a short questionnaire concerning their performance on the task. Subjects were tested in two different environments. Half of the subjects (n=28) were tested at the lab in a quiet environment, without distractions in the room (quiet testing group). The other half of subjects (n=28) performed the task in a distracting environment, such as a study room, where there are always groups of people speaking and moving around, or at a cafe with some background music and a group or two of people speaking not too loudly (noisy testing group).

These two environments (quiet and noisy) replicate common working environments. Some people need to work in absolute silence, while others like to work in a quiet cafe. Moreover, most people work in rooms with other people, so even those that prefer silence have to work with moderate environmental distractions.

In this context, our question investigates whether people subjected to distractors should consciously try to suppress their attention to the distractors and make an effort to concentrate on the task they are performing or naturally concentrate on the task they are performing, without worrying about the distractors.

EXPERIMENTAL TASK

To investigate the ironic effects of attention suppression we developed a novel experiment that evaluates the effects of attention suppression on the performance of a simple Go/NoGo task. The task was programmed in Matlab using a toolbox for neuroscience research to programme experimental tasks, the Psychophysics Toolbox Version 3 (Psychtoolbox). This toolbox guarantees accurate presentation of visual and auditory stimuli, as well as precise collection of the participants’ responses. It is very important to use appropriate software when programming this type of cognitive tasks to minimize the error in the timing of stimuli presentation or the calculation of reaction times. The toolbox used supports sub-millisecond timing.

THE GO/NOGO PARADIGM

Go/NoGo tasks are commonly used in basic research in cognitive science to study response inhibition. In these tasks, subjects are presented with a series of two possible stimuli on a screen and are required to respond (i.e., by pressing a designated key) whenever they see a Go stimulus and withhold their response (i.e., not pressing the designated key) whenever they see a NoGo stimulus.

The Go/NoGo paradigm we designed for this experiment is particularly challenging. We chose neutral images to use as stimuli in order to avoid possible associations that subjects could build based on their experience with similar symbols. The Go and the NoGo stimulus are very similar to each other (one of the stimulus is a rotation of the other stimulus by 180º – Figure 1.1) to force the processing of each stimulus before deciding to respond or not. We have previously tested this task with simpler stimuli (a green square as Go stimulus and a red square as NoGo stimulus) and observed that it made subjects' responses automatic, not requiring the cognitive processing of the stimuli that we want.

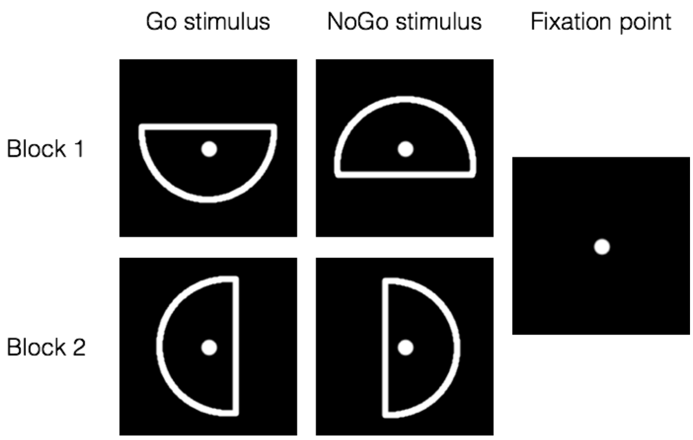


Figure 1.9 Task stimuli and fixation point. A different pair of stimuli (Go and NoGo stimulus) is used on each block of the task. The pair of stimuli used on the first or the second block was counterbalanced across the participants, as well as the stimulus that served as Go and NoGo in each pair. The fixation point is always present on the screen and the stimuli appear around the fixation.

Subjects were instructed to press the "B" key on the keyboard whenever they were presented the Go stimulus and withhold from pressing "B" when they see the NoGo stimulus. The stimuli were presented on the centre of the screen around the fixation point (Figure 1.1) and disappeared after 400 milliseconds. The subjects had to respond within those 400 milliseconds while the stimulus was on the screen. Following a 600 milliseconds inter-stimulus interval (ISI) a new stimulus was presented. Each stimulus presentation corresponded to a new trial. This task had 600 trials divided in six parts.

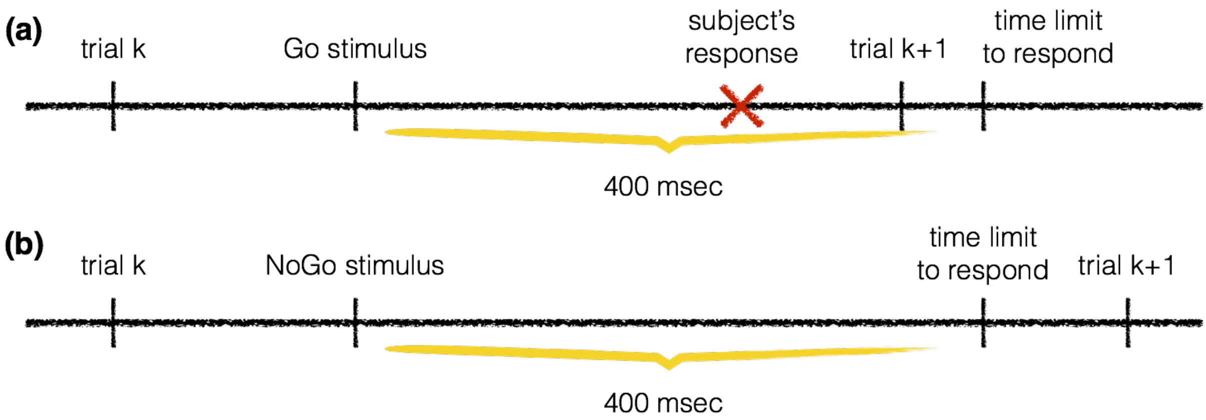


Figure 1.10 Timings of two illustrative trials on the task. Panel (a) represents a Go trial, where a Go stimulus is presented. The subject responded correctly, by pressing the response button before the time limit to respond (400 milliseconds). The next trial starts after the subject's response. Panel (b) represents a NoGo trial, where a NoGo stimulus is presented and the subject correctly inhibited his/her response. In this case, the next trial will only start after the time limit to respond has passed.

TASK DIVISION

1. Blocks – stimuli frequency division:

The task has two main blocks that differ in the frequency of the presentation of each stimulus. The high Go frequency block has 80% of Go trials and 20% of NoGo trials, whereas the low Go frequency block has 20% of Go trials and 80% of NoGo trials. Each block has 300 trials and the order of the two blocks on the task was counterbalanced across subjects.

Each block potentiates a different type of error: the high Go frequency block biases the subjects towards responding "Go" and potentiates commission errors, because subjects tend to respond even when a NoGo stimulus is presented; the low Go frequency block biases subjects towards withholding their response and potentiates omission errors, since subjects tend not to respond even if a Go stimulus is presented.

2. Condition – background images division:

Each block of the task was divided into three conditions. These conditions aimed to investigate the effects of suppressing attention from distracting images that were presented on the background of screen, around the stimuli.

No images condition:
This is the control condition where no distracting images are presented while the subjects perform the Go/NoGo task. The screen has a white background and in the centre a black square where the stimuli (Go/NoGo) are presented.

Avoid images condition:
In this condition, several abstract images are continuously presented on the background around the stimuli presentation. Subjects are instructed to avoid looking at those images and only concentrate on the central stimuli.

Allow images condition:
In this condition another group of abstract images is presented on the background around the stimuli presentation, but this time subjects are allowed to look at them if they please, while they perform the Go/NoGo task.

The order of the three image conditions was also counterbalanced across subjects.

The distracting images presented in the avoid and allow conditions were chosen to be neutral images without any specific meaning to the subjects. There were two groups of abstract images (see the appendix A.1.1 and A.1.2) and each subject was assigned a group of images to the avoid condition and the other group of images to the allow condition. This assignment was counterbalanced across subjects and will be referred to as image assignment group 1 or group 2. Image assignment group 1, means that the group 1 of images (appendix A.1.1) was assigned to the avoid condition, while the group 2 of images (appendix A.1.2) was assigned to the allow condition, and vice-versa for image assignment group 2.

The task design is described in table 1.1.

Block (Go frequency)	Condition (background images)
High Go frequency (80 % Go trials)	No images
	Avoid images
	Allow images
Low Go frequency (20 % Go trials)	No images
	Avoid images
	Allow images

Table 1.1 Ironic effects of attention suppression task design. The order of the blocks was counterbalanced across subjects, meaning that a subject could start the task either with 80% or with 20% of Go trials. On each block, the order of presentation of conditions was also counterbalanced across subjects, with the limitation that the order of presentation of the conditions in the first and second blocks was always the same.

INSTRUCTIONS

We wanted the participants to avoid looking at the avoid images not simply because they were asked to, but because they had a good reason to avoid looking at those images. To that end, we told the participants that this study aimed to investigate the effect of two groups of images on performance of the task: a group of images that helps to concentrate and another group of distractor images that worsens their performance. We instructed the participants to let the concentrating images help them naturally without having to do anything directly (allow condition) and to explicitly avoid looking at the distracting images, to prevent those images from affecting their performance (avoid condition). In the end of the study, all participants were debriefed about the true objective of this experiment.

QUESTIONNAIRE

At the end of the task, subjects responded to a short questionnaire about the task. The questionnaire had two questions, each with two sub-questions, to which subjects responded using a Likert rating scale with 7 points. All questions had to be responded and only one answer was admitted for each question.

The questions were the following:

Table 1.2 Questionnaire about the task performance.

1.	**During the task, how much did you try to look or not to look at the two types of images?**					
a) Allow images						
Really tried not to look	Tried not to look	Somewhat tried not to look	Did not try to look nor not to look	Somewhat tried to look	Tried to look	Really tried to look
b) Avoid images						
Really tried not to look	Tried not to look	Somewhat tried not to look	Did not try to look nor not to look	Somewhat tried to look	Tried to look	Really tried to look
2.	**How much did the two types of images affected your performance?**					
a) Allow images						
Worsen a lot	Worsen	Somewhat worsen	Did not help nor worsen	Somewhat helped	Helped	Helped a lot
b) Avoid images						
Worsen a lot	Worsen	Somewhat worsen	Did not help nor worsen	Somewhat helped	Helped	Helped a lot

The order of the questions (1 and 2) and the order of the sub-questions (a and b) was also counterbalanced across subjects. The questionnaire was created using the online software Survey Monkey®. The questionnaire applied to the subjects was written in Portuguese and can be found in appendix A.1.3.

DATA ANALYSIS

The experimental data was analysed using the software for statistical computing R (R version 3.1.1).

Performance on the ironic effects of attention suppression task was evaluated by exploring the reaction times of subjects' responses and the errors they made throughout the task.

Reaction times are calculated as the time interval between the presentation of the stimulus on the screen and the subject's response to that stimulus (the moment when the subject pressed the response key). Reaction times where explored for both stimuli: reaction times of responses to the Go stimulus will be referred to as Go reaction times, whereas reaction times of responses to the NoGo stimulus will be referred to as NoGo reaction times. NoGo reaction times occur when the subjects respond to a NoGo stimulus and represent an error in performance, because subjects are instructed not to press the response key whenever they see the NoGo stimulus. Nevertheless, responses to NoGo trials do occur and we chose to analyse those reaction times.

Errors in performance can be divided into two types: omission errors, that occur whenever the subjects fail to respond to a Go stimulus; and commission errors, that occur when subjects respond to a NoGo stimulus.

To test whether the subjects' performance was affected by the active suppression of the distracting images in the avoid images condition, we conducted a multivariate analysis of variance (MANOVA) with the four dependent variables described (Go mean reaction time, NoGo mean reaction time, omission errors, and commission errors). The independent variables were all categorical; there were two within-subjects factors: condition (no images, allow, avoid) and block (high Go frequency, low Go frequency), and two between-subjects factors: environmental noise (quiet testing, noisy testing) and images assignment (group 1, group 2).

The MANOVA analysis was preferred over separated ANOVA analyses to explore the subjects' task performance as a whole and not as separate responses, since we want to answer a question related to the overall performance on the task and not with specific measures of performance. Furthermore, our comparison groups may not differ at a single response, but differ jointly in the four responses (Go mean reaction time, NoGo mean reaction time, omission errors, and commission errors) and only multivariate tests are able to see those differences. On the other hand, multivariate analysis of variance accounts for the correlation between the dependent variables, which is very important in our design since our dependent variables represent repeated measures taken from the same subject and, therefore, are correlated among themselves and should be analysed as a group and not separately.

Reaction times and errors follow different distributions. While reaction times can take any positive value less than 0.4 (400 milliseconds is the limit time to respond), errors can only take two values: 0 when the subject responded correctly, or 1 if the subject made an error. Therefore, reaction times are typically normally distributed, but the proportion of errors follows a binomial distribution.

For that reason, the proportion of omission and commission errors were subjected to a logit transformation,

$$logit(p) = \ln\left(\frac{p}{1-p}\right), \tag{1}$$

where p is the probability of errors for each subject on each part of the task (condition x block). The logit transformation is widely used to transform binomial responses and transforms them in a way that, instead of being distributed between 0 and 1, they vary from negative to positive infinity and are centred on zero, which corresponds to a probability of 0.5. In logit models, differences in probabilities around p=0.5 have a lower impact than the same differences close to p=0 or p=1 (figure 1.3). Extreme probabilities of omission or commission errors, p = 1 or p = 0, were corrected to p - 1/N and p + 1/N, respectively, where N is the number of trials (Go trials when correcting omissions or NoGo trials when correcting commissions) for that condition on that block.

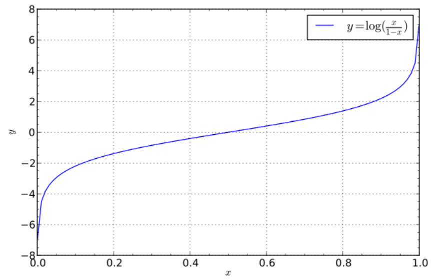


Figure 1.11 Logit curve, by Krishnavedala - Own work. Licensed under CC0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Logit.svg#/media/File:Logit.svg

STIMULUS DISCRIMINABILITY AND RESPONSE BIAS

In psychological experiments, binary responses are commonly considered to be influenced by two distinct factors: the capacity of perceptual discrimination and the response bias of the subjects. For example, considering the binary response Yes/No, subjects with similar perceptual discrimination between two stimuli can have different tendencies to respond "Yes" or "No"; if subject A is biased towards "Yes", he/she will respond "Yes" more often than subject B. On the other hand, two subjects with the same response bias can give different responses if they do not have the same perception discrimination capabilities.

To investigate perceptual discrimination and response bias, signal detection theory[41] can be applied to analyse the data from psychological experiments where the subjects have to discriminate between a specific stimulus, defined as the signal, and all other occurrences, defined as noise. In our task, the signal is a Go stimulus, whereas a NoGo stimulus can be considered noise.

Signal detection theory aims to estimate the perceptual discrimination and the subject's strategy to perform the task (the subject's response bias) from the experimental data. Response sensitivity corresponds to a parameter (d') that indicates the strength of the signal relative to the noise. Higher values of response sensitivity mean that the subject can easily discriminate between signal and noise. The subject's strategy can be evaluated through the bias (ß) to respond "Yes" or "No" (on this task "Go" or "NoGo", respectively).

These parameters (d' and ß) are calculated from the experimental data. Binary responses to signal and noise can be divided into four types of responses: responding "Yes" to signal is a hit, but responding "Yes" to noise is a false alarm; on the contrary, responding "No" to noise is a correct rejection, while responding "No" to signal is called a miss (table 1.3).

Table 1.3 Signal detection theory. Possible responses to the stimuli.

Reality	Subjects' response (decision)	
	Yes	No
Signal present	Hit	Miss
Signal absent	False alarm	Correct rejection

Subjects' responses can be characterised using only two of the four types of response described, because the two responses on each line are mutually exclusive. Herein, the responses will be characterised by the proportions of hits and false alarms.

From these data we can construct a model which assumes that the subject's response depends on the intensity of a hidden variable (e.g. evidence) that allows the subject to decide whether there is a signal or not, so that above a predefined threshold (criterion) of that hidden variable the subject will respond "Yes" and below that threshold the subject will respond "No".

The signal detection model assumes that the hidden variable generated for the noise condition can vary following a normal distribution. Furthermore, the model also assumes that the values of the hidden variable for the signal condition have the same shape as for the noise condition, but the signal distribution is added to the noise distribution, resulting in a shift of the distribution to the right ($\mu_{noise} = 0$ and $\mu_{signal} \geq 0$, σ is always equal to 1).

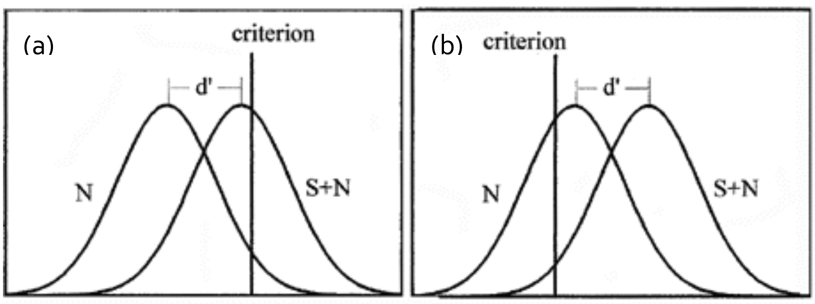


Figure 1.12 Signal detection theory. Noise (N) and signal (S+N) distributions: d' is the distance between the two distributions; greater d' values indicate that the distributions do not overlap and, therefore, there is good discriminability between signal and noise. ß is the criterion that separates positive (Go) from negative (NoGo) responses: if the value of the hidden variable is below the criterion the subject gives a negative response (NoGo), whereas for values of the hidden variable above the criterion the subject's response will be positive (Go). Consequently, higher ß values (a) indicate a tendency to give a negative response (NoGo) and lower ß values (b) indicate a tendency to respond positively (Go).
Font: http://ai.ato.ms/MITECS/Entry/macmillan.html

The distance between the two distributions is the parameter d' and is calculated as

$$d' = z(hit\ rate) - z(false\ alarm\ rate). \tag{2}$$

The threshold for the subject to decide whether there was a signal or noise is the parameter ß that is calculated as

$$\beta = e^{-\frac{z(hit\ rate)*z(hit\ rate)}{2} + \frac{z(false\ alarm\ rate)*z(false\ alarm\ rate)}{2}} \tag{3}$$

β reflects the subject's bias to respond "Yes" or "No". Unbiased subjects will have a β value around 1. The higher the bias to say "Yes", the higher the hit and false alarm rates, and β approaches zero. As the bias to say "No" increases, the hit and false alarm rates decrease, and beta increases over 1 on an open-ended scale.

Besides these two measures of discriminability and bias, other measures have been proposed. I will focus on the measures proposed by Davison and Tustin in their behavioural detection theory[42]: log d and log b, which represent response discriminability and bias, respectively. These measures are more reliable than d' and ß for analysis with fewer than 100 trials of each type, since with fewer trials d' is poorly estimated[43]. To calculate response discriminability and bias I have 20 trials of one type and 80 trials of the other type, so log d and log b where preferred, although all measures (d' and ß, log d and log b) were calculated.

$$\log d = \frac{1}{2} \cdot \log_{10}\left(\frac{hits}{misses} \cdot \frac{correct\ rejections}{false\ alarms}\right) \tag{4}$$

$$\log b = \frac{1}{2} \cdot \log_{10}\left(\frac{hits}{misses} \cdot \frac{false\ alarms}{correct\ rejections}\right) \tag{5}$$

Calculating response discriminability and bias for extreme values of hit and false alarm rates

When the proportions of hits or false alarms have extreme values (1.0 or 0.0) we cannot compute the inverse function of the distribution (z-value) and, therefore, d' and ß values cannot be calculated. Moreover, neither log d and log b values can be calculated, which would imply a division by zero. Several correction methods have been used to solve this problem. Brown and White [43] compared the different mathematical corrections typically used and showed that it is best to add a constant k to all entries of the contingency table. Their analysis indicated that the correction constant k should be between 0.25 and 0.5. In our study we chose to add a constant k=0.5 to all entries of the contingency table in order to calculate d', ß, log d and log b.

Questionnaires' responses analysis

The subjects' response to the questionnaire about the task was analysed in conjunction with their performance. Each question had seven possible responses, which were coded in a scale from -3 to 3, being 0 the neutral response and -3 and 3 respectively the extreme negative and extreme positive responses.

Responses to the questionnaire were compared to the performance of the task using correlations between the dependent variables and the responses to the questions.

RESULTS AND DISCUSSION

DEMOGRAPHIC DATA

Fifty-six subjects completed the ironic effects of attention suppression task.

Participant sample characteristics:

All participants were adults, with ages ranging from 19 to 57 years old (mean age 27.46 ± 7.77). There were 32 females and 24 males (57.14 % females).

Half of the subjects were tested in quiet conditions (quiet testing group) and the other half of the subjects were tested in a distracting environment (noisy testing group). The two groups did not differ in age ($t(53) = 1.103$, p-value = 0.275, two sample t-test), sex (X^2 (1, n=28) = 0.656, p-value = 0.418, chi-squared test) or education (X^2 (2, n=28) = 3.820, p-value = 0.148, chi-squared test). The demographic characteristics of each group of participants are depicted in table 1.4.

Table 1.4 Demographic characteristics of the participants enrolled in the experiment of ironic effects of attention suppression and divided by testing environment (quiet, noisy).

Characteristic	Quiet testing (n=28)	Noisy testing (n=28)
Age, years	28.61 ± 8.13	26.32 ± 7.36
Female	18 (64.29 %)	14 (50.00 %)
Education:		
High school graduate	0	2
Bachelor or master degree	24	25
Doctoral or higher degree	4	1

PERFORMANCE ON THE IRONIC EFFECTS OF ATTENTION SUPPRESSION TASK

The performance of the cognitive task was firstly evaluated in four main variables: the mean reaction time for Go trials, the mean reaction time for NoGo trials, proportion of omission errors and proportion of commission errors.

Hypotheses for task performance

Hypothesis 1: We were expecting to find ironic effects of trying to suppress the background images on the avoid images condition. Those ironic effects would be characterized by worse overall performance on the avoid condition, relative to the allow condition, and would be represented by:

- Higher Go mean reaction times on the avoid condition, meaning that the subjects take more time to respond to the Go stimulus.

- Higher NoGo mean reaction times on the avoid condition, indicating that the subjects' stop process is slower than in the allow condition. The stop process will not be extensively detailed since it is not the objective of this study, but I will briefly explain it so that the reaction times for the NoGo stimulus can be analysed. The traditional theory defends that the stop process competes with the go process as in a race and the faster process to reach the threshold wins. On NoGo trials, the subject should withhold the response, but that does not always happen. When the subject sees the NoGo stimulus, he/she immediately starts the go process to press the button; only after the stimulus is processed in the brain, will the stop process initiate and try to inhibit the response. If the stop process is fast enough it will win the race and the response is withhold. We hypothesised that the stop process will be slower on the avoid condition, due to ironic effects, and that would result in longer NoGo reaction times.

- Higher proportion of incorrect responses:
    - omission errors are potentiated in the low Go frequency block (20% Go), since the subject is biased towards not responding and "NoGo" is the prepotent response on this block. We expected to find bigger differences in the omission errors between the avoid and allow conditions for the low Go frequency block;
    - commission errors are potentiated in the high Go frequency block (80% Go), since the subject is biased towards responding "Go". We expected to find bigger differences in the commission errors between the avoid and allow conditions for the high Go frequency block.

Hypothesis 2: Overall performance of the task should be better for the no images condition relatively to the allow and avoid conditions, because in the no images condition the subjects are not presented with distractors around the stimuli and should be able to better concentrate on the task.

Hypothesis 3: Environmental noise should aggravate the occurrence of ironic effects, since it increases the mental load of the subjects. In fact, we even considered that the two levels of the factor environmental noise could present different results:

- For the noisy testing condition, we expected to find ironic effects that would be denoted by a worse performance for the avoid condition, relative to the allow condition.
- For the quiet testing condition, we considered that ironic effects might not occur: because the subjects will not be under high mental load, they might effectively control their mind and successfully suppress the avoid images. This can help them to perform better on the avoid condition than on the allow condition.

Mean values for each response variable were plotted. Each response is divided by the factors condition (no images, allow, avoid), block (high Go frequency, low Go frequency), and environmental noise (quiet testing, noisy testing).

MEAN REACTION TIMES:

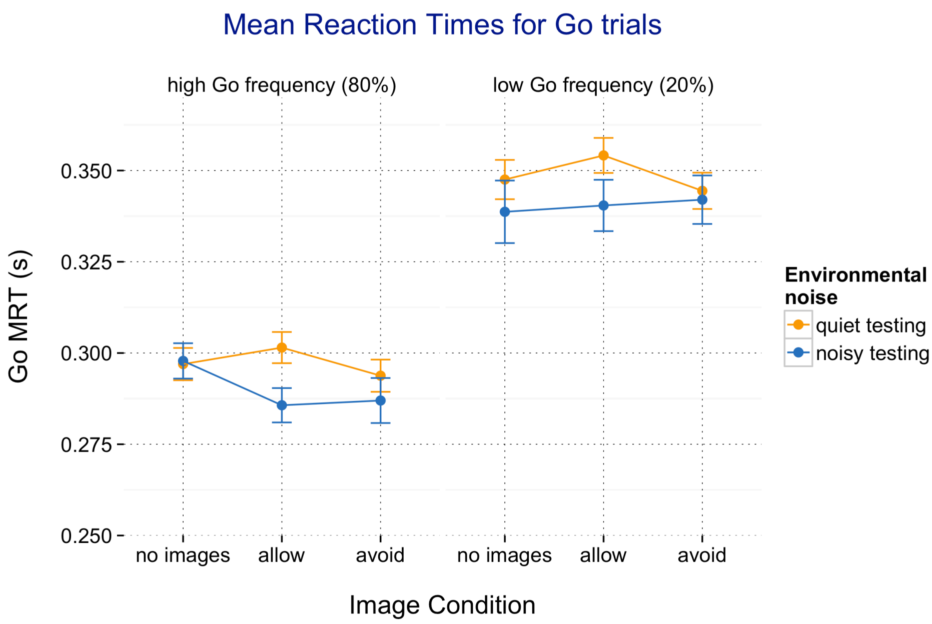


Figure 1.13 Go mean reaction times for each condition, divided by Go frequency and environmental noise. The error bars indicate the standard error of the mean.

The mean reaction times for Go trials were longer for the low Go frequency block. On this block, Go trials are infrequent and the subjects take more time to respond. On the high Go frequency block, the Go trials frequency is very high (80%) and the subjects are constantly prepared to rapidly press the response key. However, differences between the three image conditions are not pronounced, nor has the environmental noise affected the results the way we have hypothesised. In contrast, on quiet testing, there is a tendency for lower Go reaction times on the avoid condition relatively to the allow condition on both high Go frequency and low Go frequency blocks.

Mean reaction times for NoGo trials can only be registered when the subjects make a commission error (a NoGo stimulus was presented but the subjects did not withhold their responses). These are special reaction times that need a careful interpretation. A common feature of these reaction times is that they are consistently shorter than the reaction times for Go trials, which is understandable since they represent a failure in response inhibition and, therefore, only occur in trials where the Go process outperformed the stop process. Another important feature is their high variability, in relation to the reaction times for the Go trials, what is explained by the fact that they represent errors and are less frequent than Go reaction times. This results in missing values and high variability of NoGo reaction times, especially for the low Go frequency block, where commission errors are very infrequent.

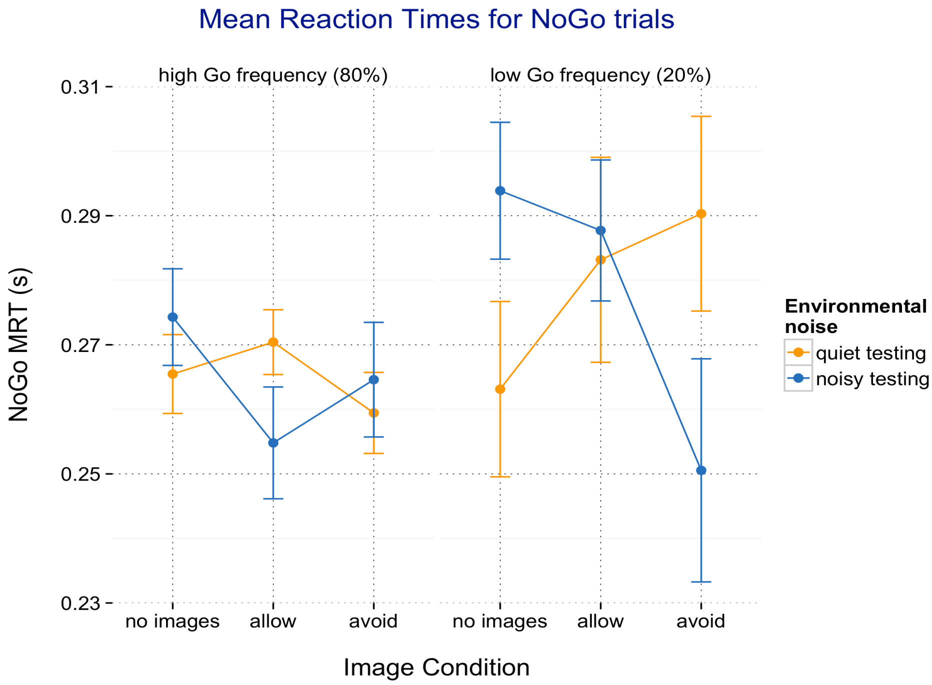


Figure 1.14 NoGo mean reaction times for each condition, divided by Go frequency and environmental noise. The error bars indicate the standard error of the mean.

Mean reaction times for NoGo trials were shorter for the high Go frequency block, because on this block the Go reaction times are shorter. Remember that only the fastest Go processes can win the race against the stop process, so on blocks where the Go reaction times are shorter the NoGo reaction times will also be shorter. The values for the two levels of environmental noise show contrary results for the allow and avoid conditions. On the high Go frequency block, NoGo reaction times were longer for the allow condition on quiet testing and longer for the avoid condition on noisy testing, as we have hypothesised. Nonetheless, on the low Go frequency block, the results are the opposite, with worse performance for the avoid condition on quiet testing and for the allow condition on noisy testing. NoGo reaction times on the low Go frequency block are highly infrequent and should not be interpreted, since they represent an exception that is not really informative in this task.

Incorrect responses:

Figure 1.15 Proportion of omission errors for each image condition divided by Go frequency and environmental noise. The error bars represent the standard error of the mean.

As expected, the proportion of omission errors was higher in the low Go frequency block, where the Go stimulus is a rare event (20%) and subjects are biased towards not responding (NoGo is the prepotent response). Omission errors tend to be higher on the allow condition in quiet testing and on the avoid condition in noisy testing, what is in line with our predictions.

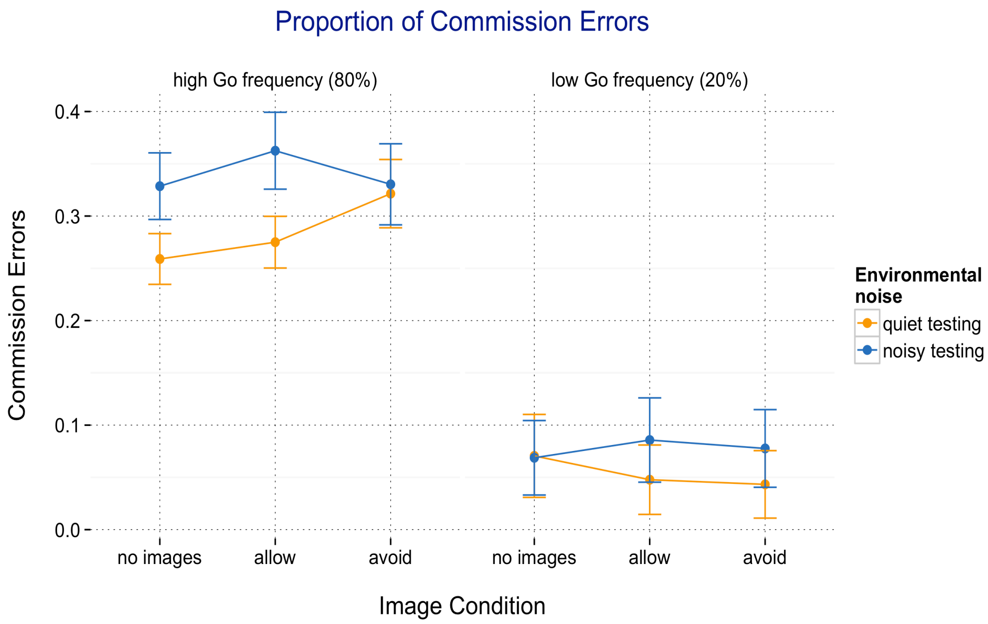


Figure 1.8 Proportion of commission errors for each image condition divided by Go frequency and environmental noise. The error bars represent the standard error of the mean.

The commission errors have an opposite behaviour in relation to omission errors: while commission errors occur more often in the high Go frequency block, where subjects are biased towards responding "Go", omission errors are more prevalent in the low Go frequency block, where the subjects are biased towards responding "NoGo". Differences for commission errors between the allow and avoid conditions can be found for the high Go frequency block, where the proportion of commission errors was higher for the avoid condition in quiet testing, and the allow condition in noisy testing. This finding seems to contradict our main hypothesis; however, commission errors are always correlated with omission errors: when a subject makes more omission errors, he/she will, as a consequence, make less commission errors and vice-versa. Therefore, commission errors between the allow and avoid conditions have the inverse result relatively to omission errors.

To investigate the effect of images condition (no images / allow / avoid) and Go stimulus frequency (80% / 20%) on the performance of the task, we performed a mixed model multivariate analysis of variance (mixed MANOVA), with dependent variables mean reaction time for Go trials, mean reaction time for NoGo trials, proportion of omission errors and proportion of commission errors, and independent variables the within-subjects factors condition (no images, avoid and allow) and block (high Go frequency and low Go frequency) and the between-subjects factors environmental noise (quiet and noisy) and images assignment (group 1 or group 2). Omission and commission errors were not normally distributed and a logit transformation was applied to both variables.

The test statistic used in the MANOVA was Pillai's criterion and a MANOVA with sum of squares type I was performed. The MANOVA analysis indicated a significant multivariate effect for the four response variables as a group in relation to the block of the task ($F(1,4) = 21.4613$, p-value = 3.491×10^{-15}), and the three-way interaction between condition*block*noise ($F(2,8) = 1.9988$, p-value = 0.04487). However, the main effects of condition ($F(2,8) = 0.3409$, p-value = 0.94978) and noise ($F(1,4) = 1.6494$, p-value = 0.16251) were not significant.

The highly significant main effect of block was expected, since there is a big difference in the frequency of the Go stimulus in the two blocks, implying that the participants will have a natural tendency to respond "Go" on the high Go frequency block and respond "NoGo" on the low Go frequency block. Consequently, reaction times will be faster on the high Go frequency block (both reaction times for Go and NoGo trials), omission errors will be infrequent, whereas commission errors will be enhanced.

The Royston's multivariate normality test indicated that our data is not multivariate normal (p-value = 3.148×10^{-11}) and a multivariate Quantile-Quantile plot for normality shows exactly the same (figure 1.9). Nonetheless, the Mauchly test for sphericity was not violated and it is perfectly reasonable to perform the MANOVA analysis with our data. In fact, there is no viable alternative multivariate test to overcome this problem, since the – theoretically – more appropriate non-parametric tests, like the Kruskal-Wallis test, can only be applied to simpler models including only one factor and cannot deal with both between-subjects and within-subjects factors.

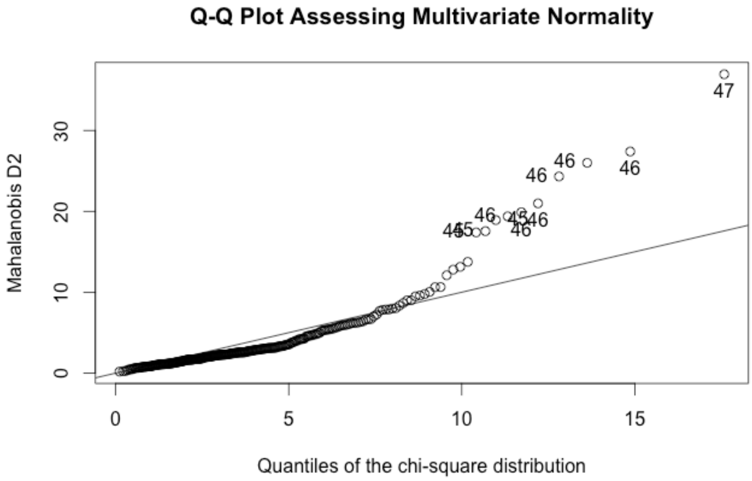


Figure 1.9 Quantilie-Quantile plot to assess multivariate normality of the data. The points at right end of the pattern are above the line, indicating a long right tail of the data distribution.

Visualizing the results from the MANOVA is not simple, since we have four dimensions (four dependent variables) divided by four independent factors. We used hypothesis-error plots (HE plots) to represent the effects of the independent factors on the response variables. However, these plots are only implemented with within-subjects effects and, due to this technical limitation, we cannot visualise these effects considering block and condition as repeated measures on the same subject.

Hypothesis-error plots use ellipses to represent hypothesis and error sums of squares and product matrices. We choose to selectively plot the three-way interaction of condition x block x noise, so that the results can be readable; more complete hypothesis-error plots can be found in appendix A.1.5.

In hypothesis-error plots, the error ellipse represents the data ellipse for the residuals, whereas the hypothesis ellipse represents the data ellipse of fitted values under the alternative hypothesis. Significant alternative hypotheses should protrude from the error ellipse, however it does not happen with the three-way interaction for condition x block x noise (figure 1.10). That can be due to the fact that the hypothesis-error plots did not consider the factors condition and block as repeated measures and, therefore, show non-significant results. Additionally, we can observe from the analysis of the graphs that the response variables that most influenced the results of the MANOVA were the mean reaction times for NoGo and the proportion of commission errors. We can also observe the distribution of the means for each condition, represented with the black points with labels 0, 1 and 2, that correspond to the grand mean for the conditions no images, avoid, and allow, respectively. Panels (a), (c), and (d) reveal that there is a tendency for the no images and the avoid conditions' mean to be further apart, with the allow condition having an intermediate value. This tendency is not significant, as we already have concluded by the non-significant main effect of condition in the results of the MANOVA, however it is a consistent tendency in our data and is also coherent with our main hypothesis for this task that the no images condition would have the best performance, the avoid condition would reveal the worst performance (due to the occurrence of ironic effects), and the allow condition should have an intermediate effect.

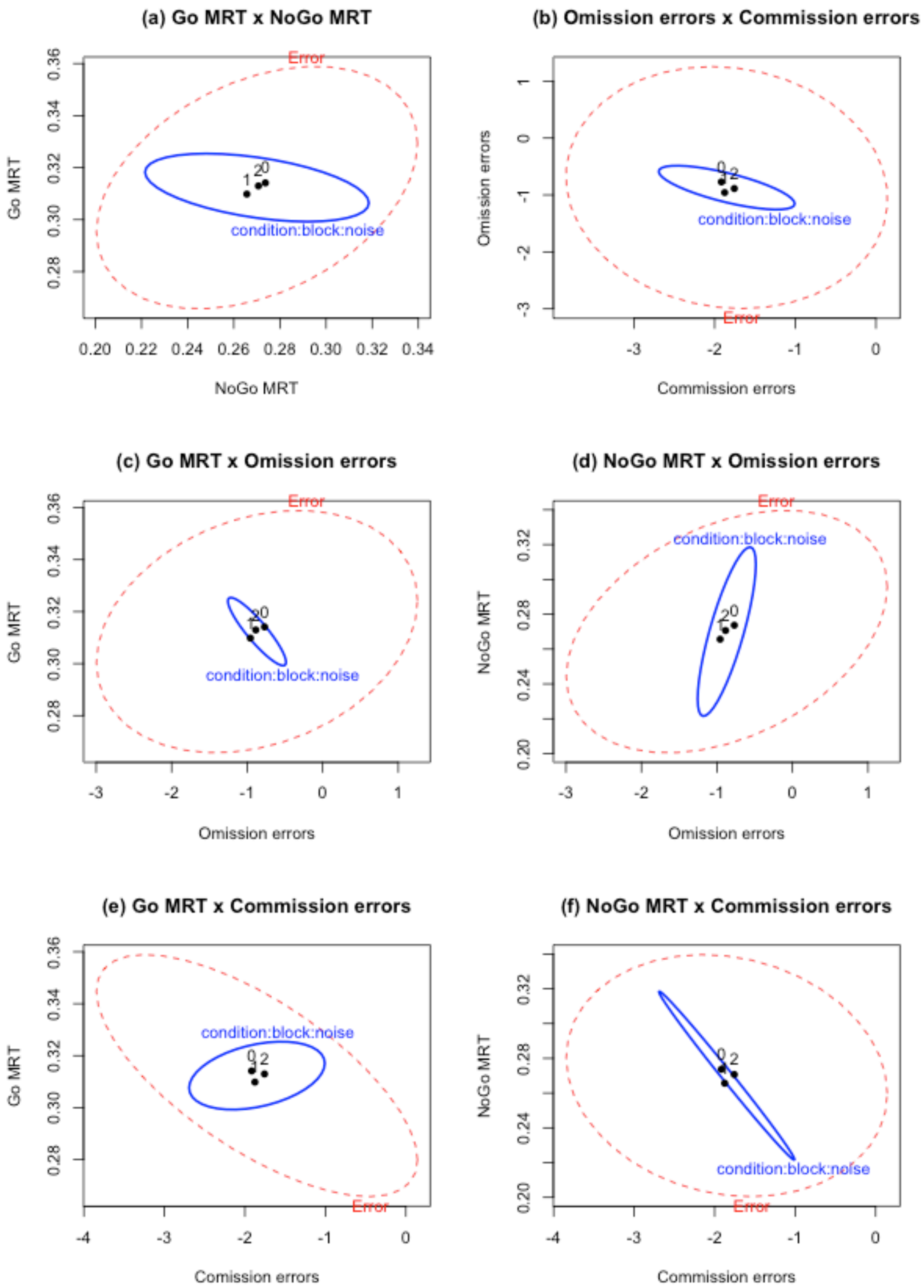


Figure 1.16 Hypothesis-error plots for the three-way interaction condition x block x noise.

The significant three-way interaction condition x block x noise is an important result, but is not easy to interpret because the factor condition has three levels (no images, allow, and avoid) and we do not know whether there was a significant difference among the three or only between two conditions. In order to clarify that question, independent MANOVA analyses were conducted for the different combinations of the three levels of the factor condition:

1. The MANOVA between the conditions No Images and Allow indicated a main effect of block ($F(1,4) = 14.8668$, p-value = 2.278×10^{-10}), but no other main effect was significant nor any interaction.
2. The MANOVA between the conditions No Images and Avoid indicated a main effect of block ($F(1,4) = 12.8149$, p-value = 4.299×10^{-09}) and a significant three-way interaction of condition x block x noise ($F(1,4) = 3.0987$, p-value = 0.01721).
3. The MANOVA between the conditions Allow and Avoid also indicated a main effect of block ($F(1,4) = 15.4144$, p-value = 1.1×10^{-10}). The main effect of noise ($F(1,4) = 2.1553$, p-value = 0.07645), and the three way interaction condition x block x noise ($F(1,4) = 2.0298$, p-value = 0.09271) were nearly significant.

The performance for the avoid condition is the one that most contributes to the significant interaction condition x block x noise. The separated MANOVAs indicated that performance on the avoid condition is significantly different from the performance on the no images condition, and the allow condition is somehow in the middle.

To better visualise these effects, plots for within-subjects analyses are presented. In these plots, the individual values for each subject were subtracted between the different conditions (allow - no images; avoid - allow; and avoid - no images) before calculating the group mean value. The results from the within-subjects analyses were compared with t-test analyses, but, although there might be some tendencies in the data, none effect was significantly different from zero.

Mean reaction time for Go trials had a negative value in the condition avoid-allow for quiet testing, suggesting that subjects had longer reaction times in the allow condition than in the avoid condition. This is congruent to our hypothesis that subjects in the quiet testing group could have better performance on the avoid condition, relatively to the allow condition.

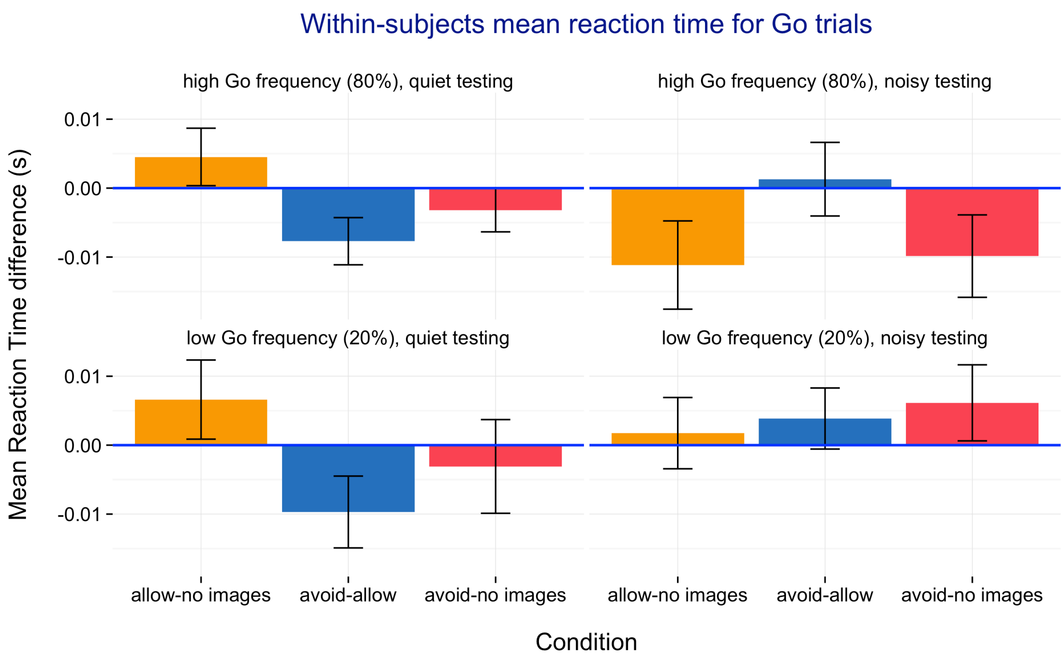


Figure 1.17 Mean within-subjects differences for Go reaction times between the levels of condition. The error bars represent the standard error of the mean.

Mean reaction time for NoGo trials between the avoid and allow condition was practically zero. There is a negative tendency for the mean value of avoid-allow on the low Go frequency block of noisy testing, indicating shorter NoGo reaction times for the avoid condition.

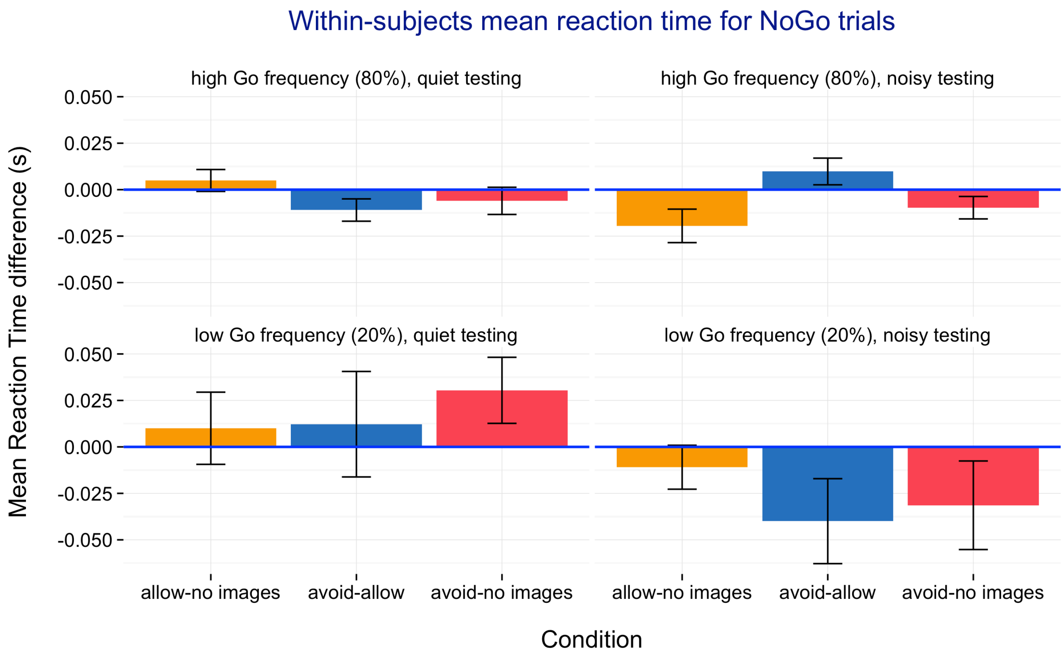


Figure 1.18 Mean within-subjects differences for NoGo reaction times between the levels of condition. The error bars represent the standard error of the mean.

Within-subjects mean omission errors tend to be negative for avoid-allow in the quiet condition, indicating more omissions for the allow condition, and positive for avoid-allow in the noisy condition, indicating more omissions for the avoid condition. Although not significant, this tendency can suggest the presence of ironic effects on the noisy testing group, and an efficient suppression of the distracting images in the avoid condition on the quiet testing group.

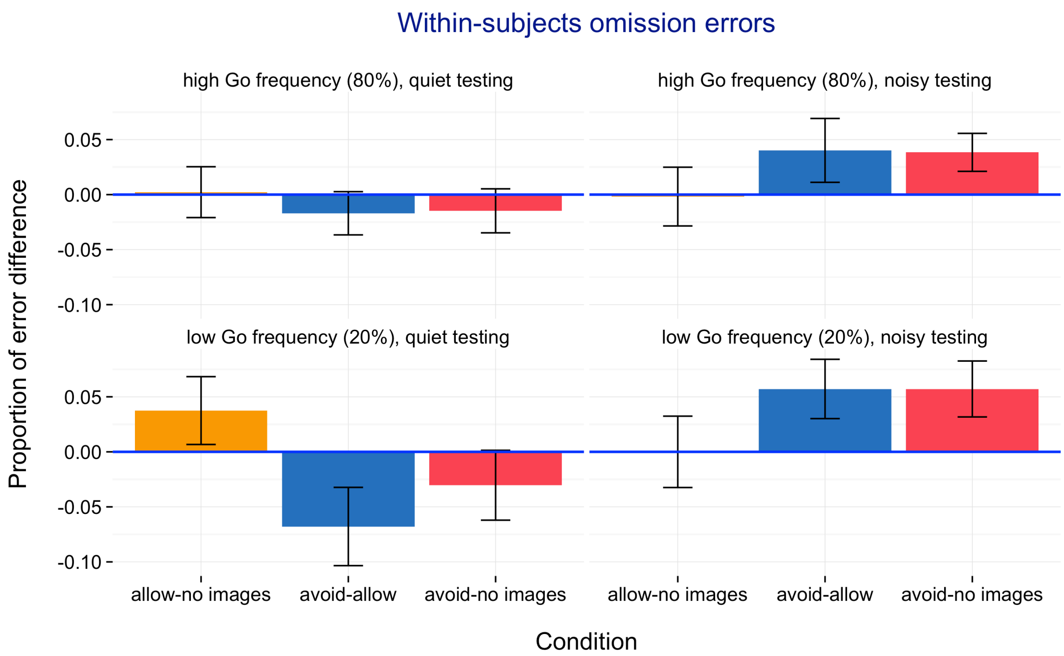


Figure 1.19 Mean within-subjects differences for omission errors between the levels of condition. The error bars represent the standard error of the mean.

Commission errors for the low Go frequency block are very small, however for the high Go frequency block there is a positive value for avoid-allow on the quiet testing group, indicating more commission errors for the avoid condition, and a negative value for avoid-allow on the noisy testing group, indicating more commissions in the allow condition.

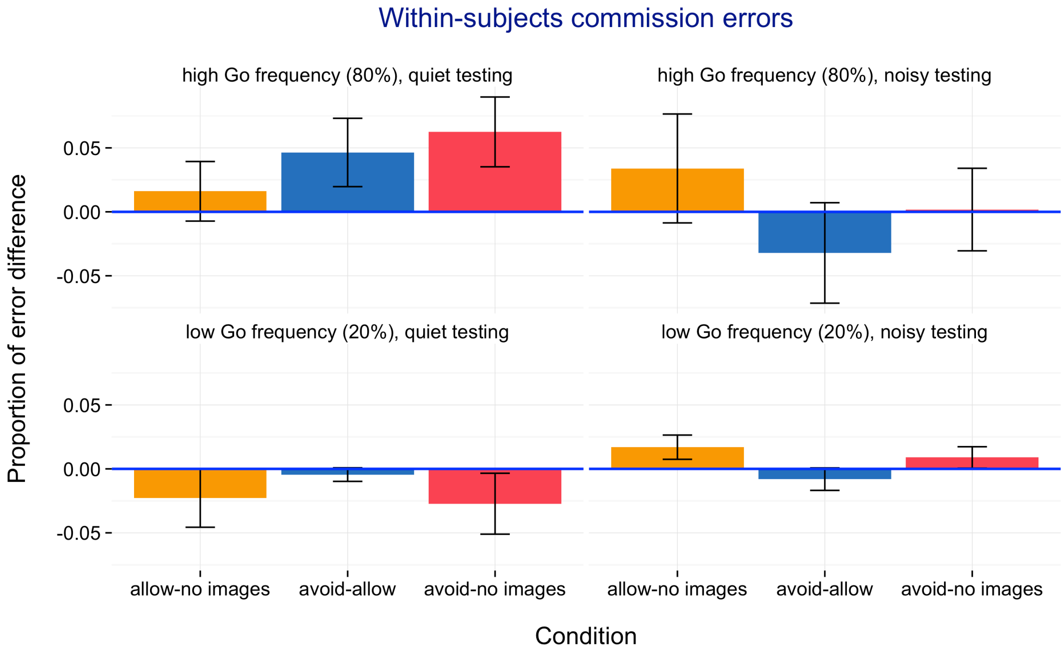


Figure 1.20 Mean within-subjects differences for commission errors between the levels of condition. The error bars represent the standard error of the mean.

To complement our attempt to understand which image conditions contributed the most to the significant results on the task, a generalised canonical discriminant analysis for the factor condition was performed. This analysis computes canonical scores and vectors that represent a transformation of the original variables into a canonical space of maximal differences for that factor, controlling for the other terms of the model (block, noise, and image assignment). The results determine the number of significant canonical dimensions, where all canonical variates are mutually uncorrelated. For our factor condition (three levels = three dimensions), the canonical discriminate analysis has reduced one dimension, resulting in two canonical dimensions (Can1 and Can2) that fully explain the variance of the data.

The results for the first canonical dimension are shown in figure 1.15. We can see that this canonical dimension explains the majority of the data variability (77%) for the factor condition. A hypothesis-error plot for the two calculated canonical dimensions can better show the dispersion of the results for each level of condition. The hypothesis ellipse for condition is plotted in canonical discriminant space and provides a low-rank of the effects for the levels of condition in the space of maximum discrimination. Here, we can see again the grand mean for the avoid condition more separated from the grand means of the no images and the allow condition.

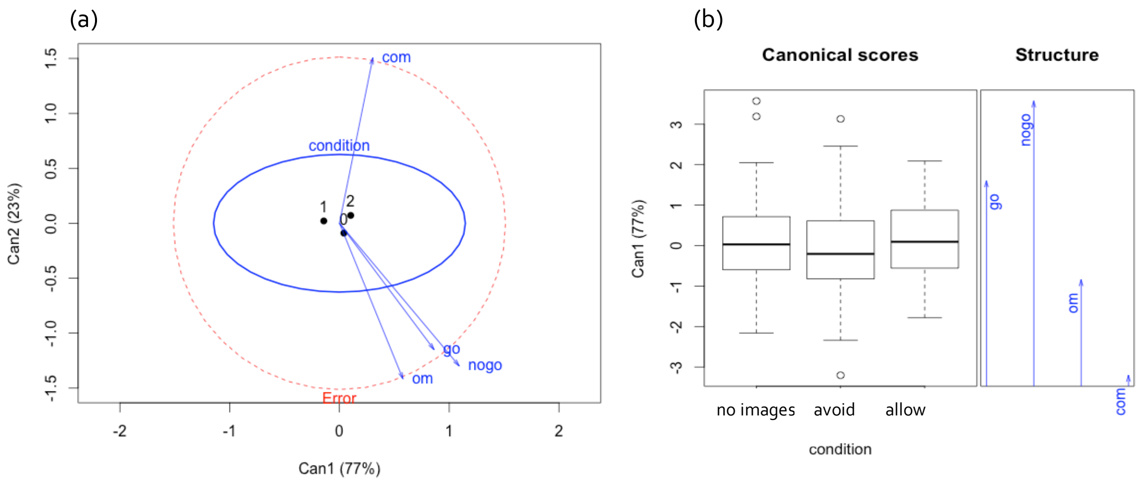


Figure 1.21 Canonical discriminant analysis results for the factor condition. (a) The hypothesis-error plot represents the error ellipse and the hypothesis ellipse for condition; the vectors represent the variation of the dependent variables on the two canonical dimensions. (b) The distribution of the data for each level of condition is represented for the first canonical dimension Can1 in separated boxplots. The vectors indicate the structure of the first canonical dimension for the four dependent variables. (go = Go mean reaction times; nogo = NoGo mean reaction times; om = omission errors; com = commission errors)

RESPONSE DISCRIMINABILITY AND BIAS

To better study the subjects' responses, measures of stimuli discriminability and response bias were calculated. Both measures from signal detection theory (d' and ß) and from behavioural detection theory (log d and log b) were calculated, since the former are more commonly used but the latter are more accurate for smaller samples (N<100).

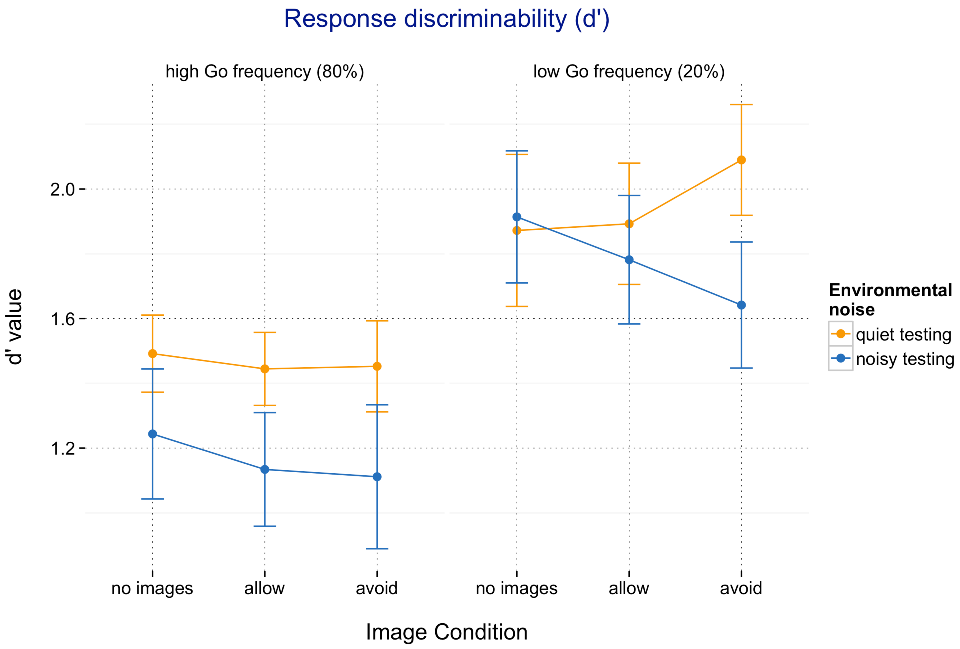
Hypotheses for response discriminability and bias

Hypothesis 1: Measures of response bias should indicate a tendency towards responding "Go" on the high Go frequency block (indicated by a ß value near zero and by a positive log b value), and responding "NoGo" on the low Go frequency block (positive ß value and negative log b value).

Hypothesis 2: Measures of discriminability might not show big differences since the stimuli used are always the same. There might be lower discriminability (lower d' and log d values) for conditions where the performance is expected to be worse: on the allow condition in quiet testing and the avoid condition in noisy testing.

Results for response discriminability and bias

Response discriminability varied for the two blocks of the task, with both d' and log d measures indicating a higher discriminability for the low Go frequency block. Also, the quiet testing group had consistent higher discriminability than the noisy testing group, indicating that in quiet conditions the subjects were able to better discriminate between the Go and NoGo stimuli and suggesting that it was harder for the noisy testing group to perform this task. Our predictions were met, but only for the low Go frequency block, where indeed discriminability was lower for the allow condition in quiet testing and the avoid condition in noisy testing.



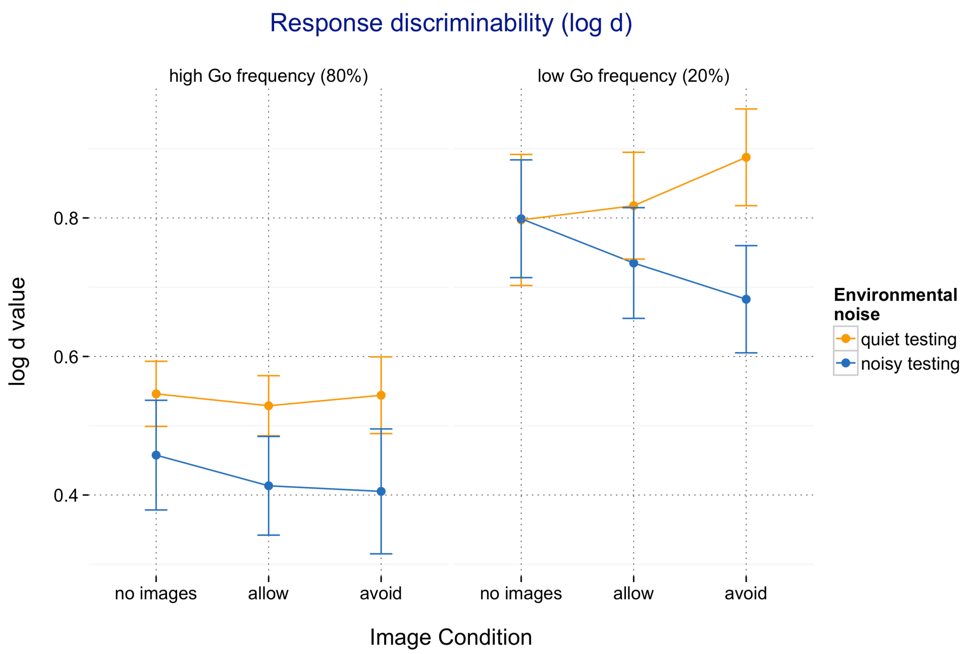
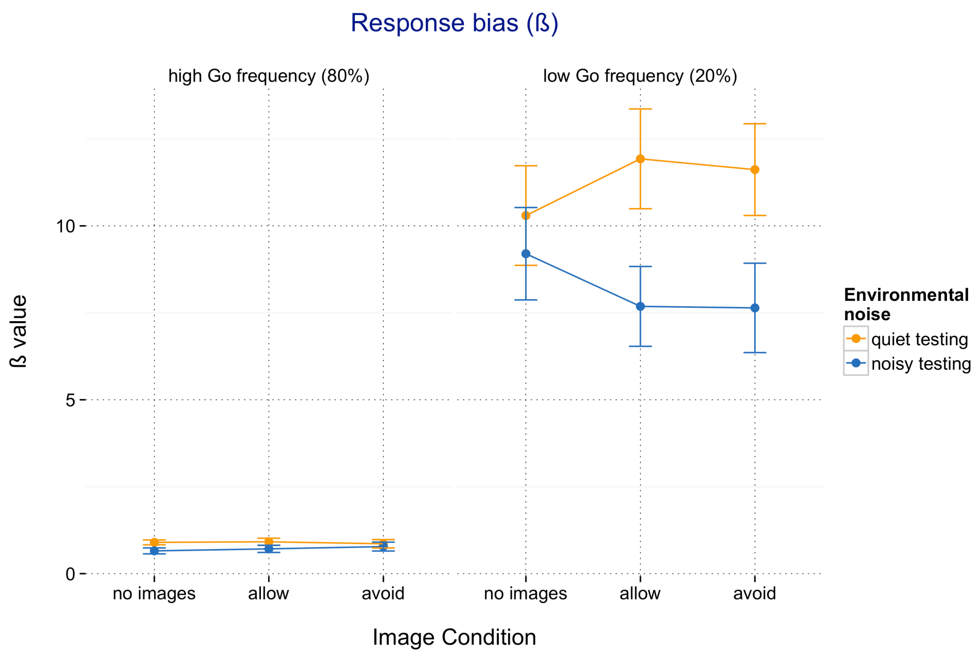


Figure 1.22 Measures of stimulus discriminability: mean values for d' and log d divided by image condition, Go frequency, and environmental noise.

As expected, response bias towards responding "Go" was found for the high Go frequency block (ß values near zero and positive log b values), while bias toward "NoGo" was found for the low Go frequency block (positive values for ß and negative values for log b), and here more pronounced for the quiet testing group (for all image conditions).



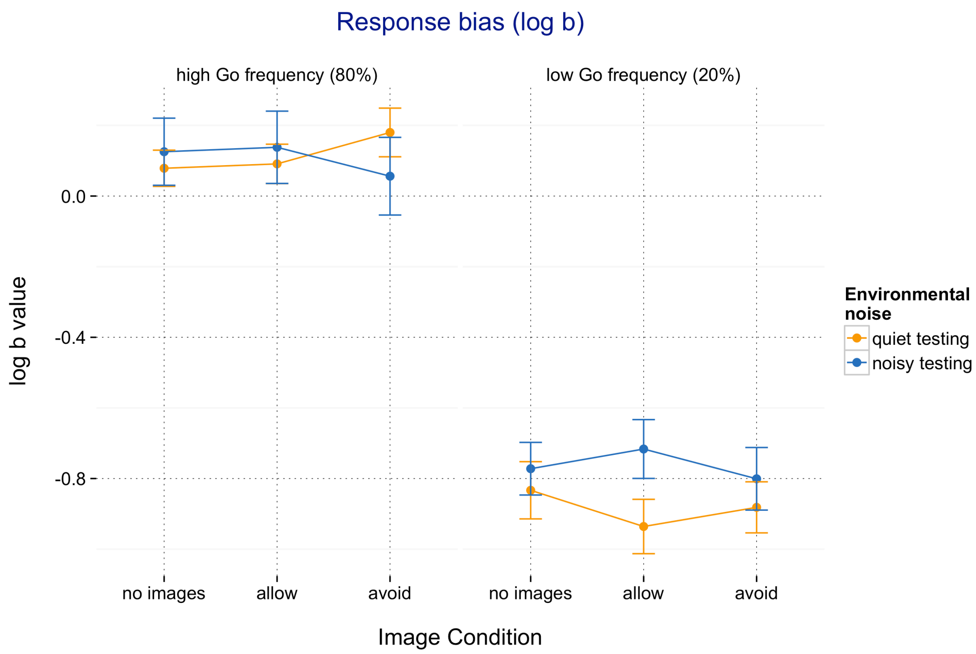


Figure 1.23 Measures of response bias: mean values for ß and log b divided by image condition, Go frequency, and environmental noise.

To confirm that the measures of response discriminability and bias are good representations of the errors in performance (omissions and commissions), a MANOVA with dependent variables Go MRTs, NoGo MRTs, log d and log b, with factors condition, block, noise and image assignment was conducted. A main effect of block ($F(1,1) = 22.2826$, p-value = 1.07×10^{-15}) and a significant three-way interaction condition x block x noise ($F(2,8) = 1.9994$, p-value = 0.04479) were shown. These results are very similar to the previous results from the MANOVA with omission and commission errors. That is because the errors are closely related to the measures of discriminability and bias; therefore, we suggest that it is equivalent to directly use the proportion of errors or these measures of discriminability and bias in our dataset.

IRONIC REBOUND EFFECTS ANALYSIS

The literature of ironic effects of mental control frequently reports rebound effects, that occur in the time period after suppression, stronger than the actual ironic effects occurred during the suppression phase. We decided to conduct an analysis to evaluate if rebound effects were significant on the performance of this task.

Herein, rebound effects can be seen when the subject performs the allow condition (analogue of the expression phase) following the avoid condition (analogue of the suppression phase).

In order to do that, a factor named rebound was introduced in the analysis. The factor rebound has two levels: "allow after avoid", representing the allow conditions that were performed after the avoid condition, and "other allow", which includes allow conditions that were performed in the beginning of the task or after a no images condition. Because the design of the task is counterbalanced, not all subjects have the same order for presentation of the conditions, and the levels of the factor rebound define two groups of subjects: there were 18 subjects on the "allow after avoid" level and 38 subjects on the "other allow" level.

Hypothesis for ironic rebound effects:

Hypothesis 1: The performance on allow conditions that were completed after an avoid condition should be worse than performance on other allow conditions, suggesting the occurrence of ironic rebound effects. This prediction is based on previous rebound effects found in other studies, which indicate that after active suppression (herein, suppression of the background images), the suppressed stimulus becomes more relevant and intrudes more often in our thoughts or, as the case of this task, in our attention.

Results for ironic rebound effects:

A MANOVA analysis was performed with dependent variables Go mean reaction times, NoGo mean reaction times, omission errors, and commission errors, and independent factors rebound, block, noise, and image assignment. A main effect of block ($F(1,4) = 9.0609$, p-value = 5.885×10^{-06}) was shown, as well as a significant three-way interaction of rebound x block x image assignment ($F(1,4) = 4.6222$, p-value = 0.0023). Other factors revealed an important effect, although did not reach significance: the main effect of noise ($F(1,4) = 2.4752$, p-value = 0.052052), the three-way interaction block x noise x image assignment ($F(1,4) = 2.1667$, p-value = 0.081665), and the four-way interaction rebound x block x noise x image assignment ($F(1,4) = 2.3593$, p-value = 0.061668).

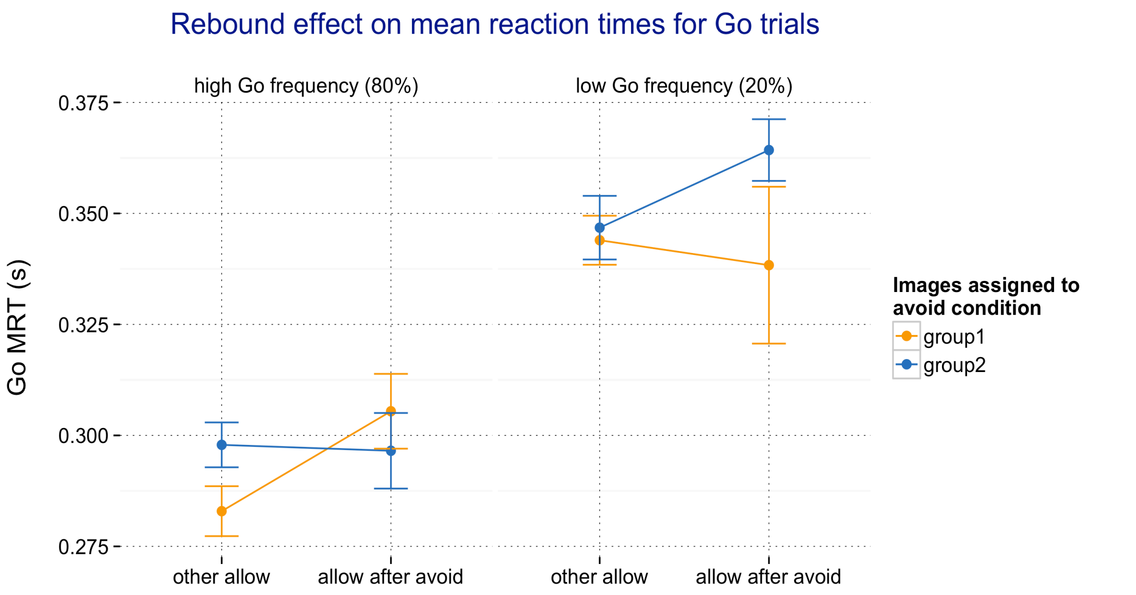


Figure 1.24 Ironic rebound effects for Go mean reaction time. The error bars represent the standard error of the mean.

Rebound effects seem to have a contradictory tendency for the two levels of images assignment. For Go mean reaction times, we can see a tendency for the occurrence of rebound effects on the images assignment group 1 in the high Go frequency block and on the images assignment group 2 in the low Go frequency block. The same tendencies can be seen for the NoGo mean reaction times.

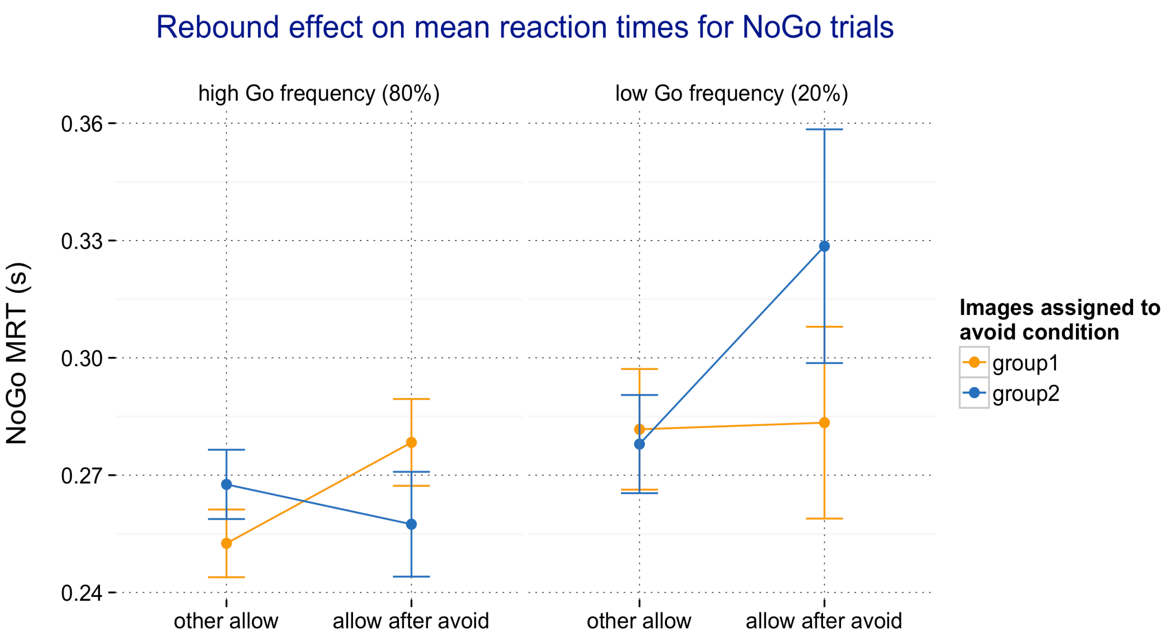


Figure 1.25 Ironic rebound effects for NoGo mean reaction time. The error bars represent the standard error of the mean.

Incorrect responses show the opposite tendency for ironic rebound effects, with higher proportion of errors on the allow after avoid level for the images assignment group 1 on the low Go frequency block (both for omission and commission errors). Omission errors also show an increase for the allow after avoid level on the high Go frequency block, but no tendency for ironic rebound effects is shown for commission errors on this block.

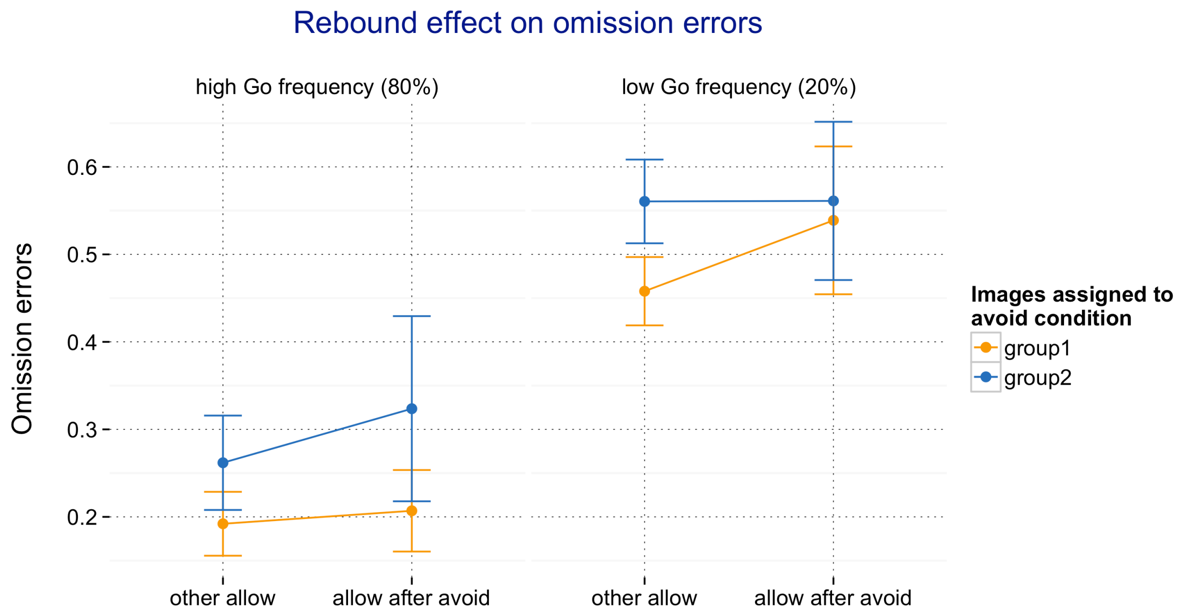


Figure 1.26 Ironic rebound effects for omission errors. The error bars represent the standard error of the mean.

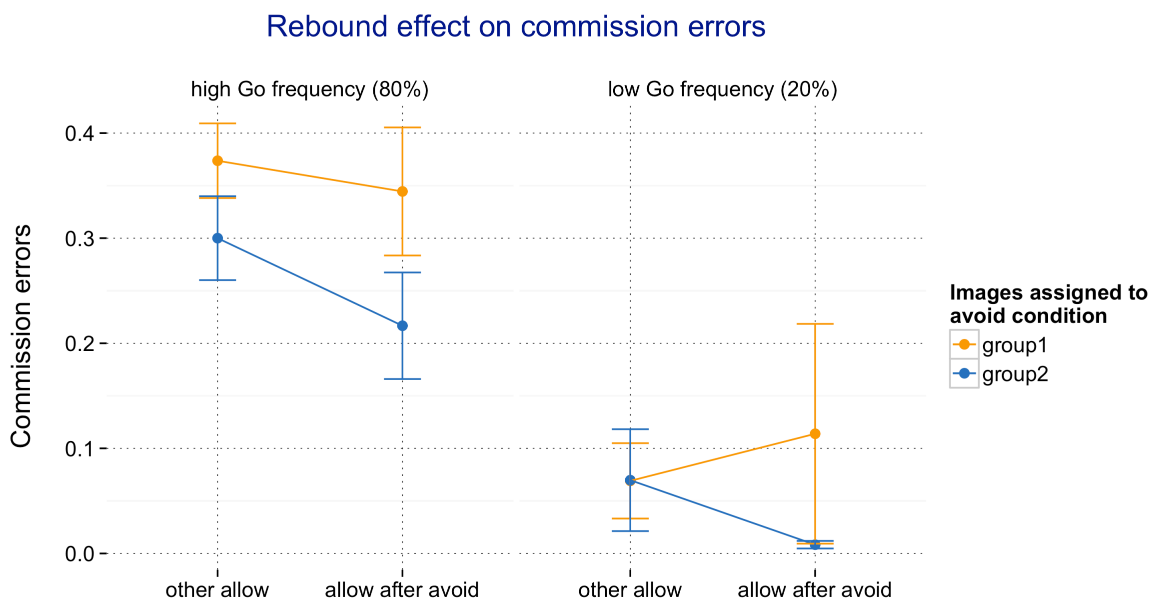


Figure 1.27 Ironic rebound effects for commission errors. The error bars represent the standard error of the mean.

Image assignment is a factor that indicates the group of images that were presented on the avoid condition. If group 1 of images was assigned to the avoid condition, group 2 of images was assigned to the allow condition for that subject, and vice-versa. Both groups of images are abstract images that should not affect preferentially the performance. The factor images assignment was introduced in the analysis to control for any variability on the data that could be a result of it, however it was not expected to be a relevant factor to explain the results. The analysis of ironic rebound effects demonstrates important results in our experiment, however it's interpretation is not straightfoward and it is difficult to understand whether this significant result indicates the occurrence of ironic rebound effects or if it is due to the unbalanced number of subjects in the two levels of rebound (other allow vs. allow after avoid).

QUESTIONNAIRE'S RESPONSES

The responses to the four questions of the questionnaire were coded in a scale with seven levels, from -3 to 3. The first two questions asked whether participants tried or not to look at the allow and the avoid images, respectively. The available responses and respective coding values were the following:

Response	Really tried not to look	Tried not to look	Somewhat tried not to look	Did not tried to look nor not to look	Somewhat tried to look	Tried to look	Really tried to look
Value	-3	-2	-1	0	1	2	3

The subjects were instructed not to look at the avoid images, so we expected them to provide a negative response to that question. More importantly, we expected that the subjects had tried harder to not look at the avoid images than to the allow images. To test this, the difference between the response values for the allow condition - avoid condition were calculated for each subject.

The last two questions asked the participants how much did they felt that the avoid and allow images affected their performance on the task. The available responses and respective coding values were the following:

Response	Worsen a lot	Worsen	Somewhat worsen	Did not helped nor worsen	Somewhat helped	Helped	Helped a lot
Value	-3	-2	-1	0	1	2	3

The mean values for each response and the calculated difference between them are presented in figure 1.22. Subjects reported to have tried harder to not look at the avoid images than to the allow images, nonetheless the difference between the two was small and the subjects tend to suppress both the avoid and the allow images. In the end, subjects reported that the allow images effect on their performance was null, but avoid images affected negatively their performance; the difference is again positive but small.

This questionnaire revealed that the subjects did not rigorously follow the instructions and a big proportion of the subjects did not believe in our story that the allow images were there to help them concentrate and the avoid images were there to distract them.

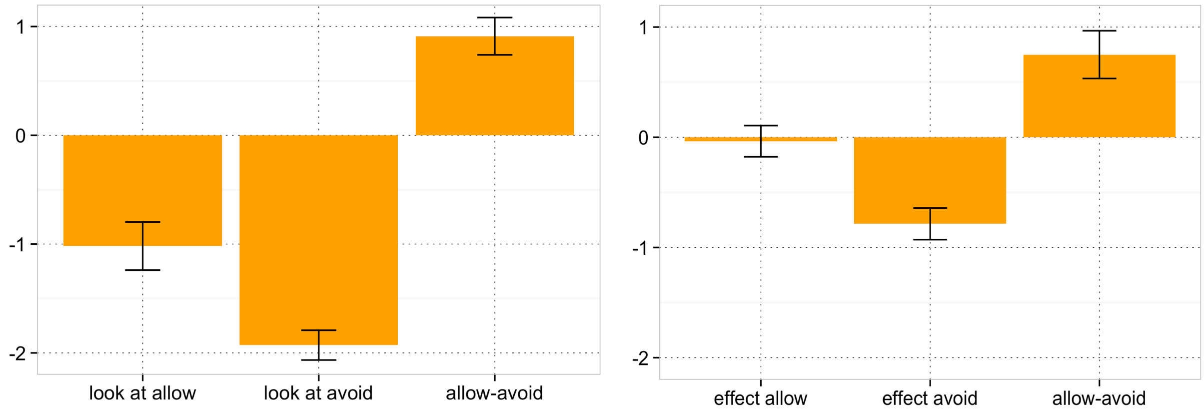


Figure 1.28 Subjects mean responses in the questionnaire about the task performance.

Correlation analyses were conducted to understand how the responses in the questionnaire relate to the performance of the task. The results are plotted in correlograms, where the blue colour represents a positive correlation and the red colour represents a negative correlation. The strength of the correlation is indicated by the colour saturation and by the size of the filled slice on the pie charts on the upper diagonal.

Two correlation analyses were conducted:

1. To compare the responses to the questions on the questionnaire that were relative to the allow condition with the performance of the task on the allow condition (figure 1.23).
2. To compare the responses to the questions on the questionnaire that were relative to the avoid condition with the performance on the avoid condition (figure 1.24).

The strongest correlations are seen within the response variables of the task, and not with the responses to the questionnaire (the latter were not significant for both allow and avoid conditions). Nonetheless, the response to whether the subjects had tried or not to look at the allow images correlated positively with the mean reaction time for Go trials and the omission errors, and negatively with the commission errors, although those results are not significant when correcting for multiple comparisons. This relationship can be interpreted in two ways: subjects that tried harder to look at the allow conditions had longer Go reaction times, more omission errors, and less commission errors; or subjects that tried harder to not look at the allow images, had faster Go reaction times, less omission errors, and more commission errors. The latter deduction is more plausible, since the questionnaire responses revealed that, in average, the subjects did not try to look at the allow images. This can indicate that trying to suppress their attention to the images help the subjects to better concentrate on the task.

CORRELOGRAM: CORRELATION BETWEEN THE QUESTIONNAIRE'S RESPONSES AND THE PERFORMANCE ON THE ALLOW CONDITION

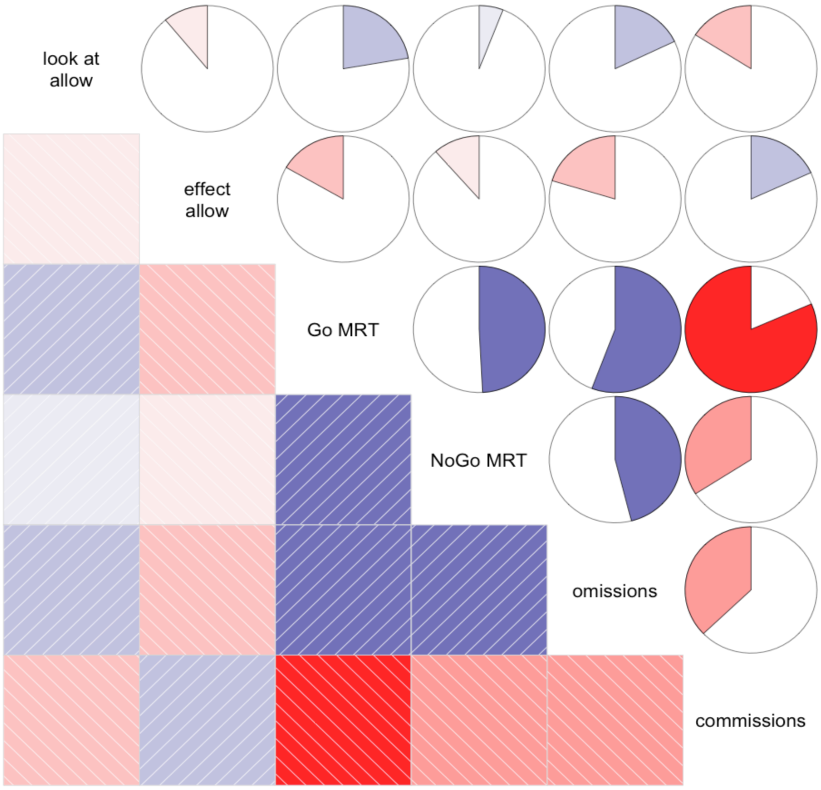


Figure 1.29 Correlogram depicting the correlation between the responses to the questionnaire relative to the allow condition and the subject's performance on the allow condition.

The response for the effect that the allow conditions had on the performance of the task correlated negatively with the Go mean reaction times and the omission errors, and positively with the commission errors. However, none of this was significant.

The strong correlations between the four dependent variables directly obtained from the participant's performance reinforce the previous results discussed: the Go mean reaction time correlates positively with the NoGo mean reaction time, and with the omission errors, but has a strong negative correlation with the commission errors; NoGo mean reaction times correlate negatively with commission errors, since subjects make more commission errors, when they respond faster, and vice-versa; furthermore, the omission and commission errors are negatively correlated.

The correlations for the condition avoid present a different pattern. An important finding is that the responses to both questions have a positive correlation, meaning that the subjects who have tried harder to avoid the images felt a higher negative effect of those images on their performance. This result might be because of ironic effects, where those who suppressed the images more intensely, experienced more ironic effects and had a worse performance. Alternatively, those subjects could be more susceptible for a suggestion effect, so that, when we told them those images would affect negatively their performance and that they have to avoid looking at them, they followed faithfully our instructions and felt that the images affected their performance just because we told them so.

CORRELOGRAM: CORRELATION BETWEEN THE QUESTIONNAIRE'S RESPONSES AND THE PERFORMANCE ON THE AVOID CONDITION

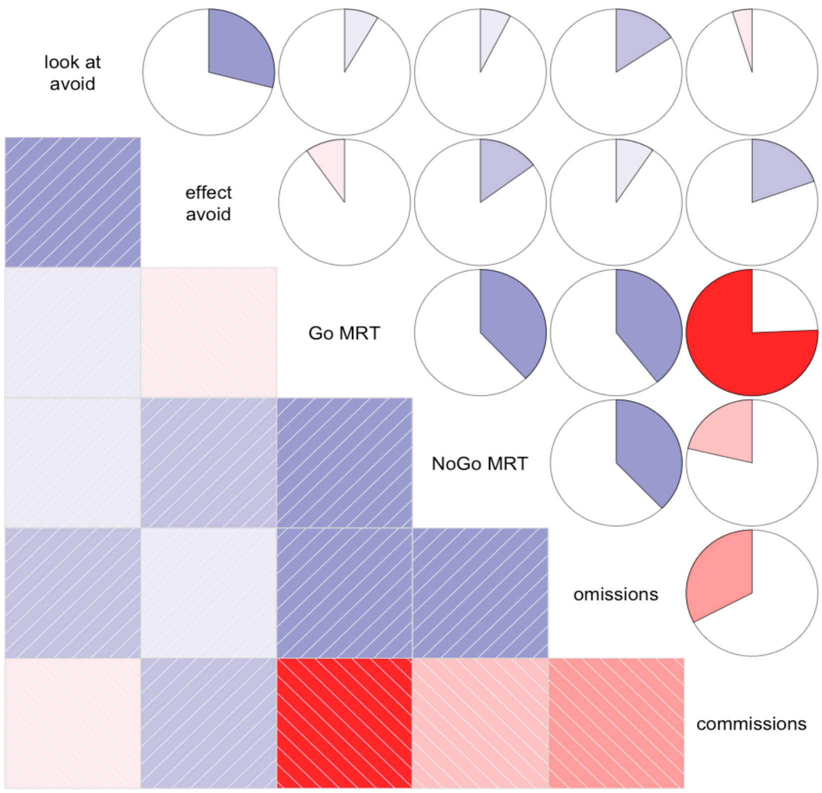


Figure 1.30 Correlogram depicting the correlation between the responses to the questionnaire relative to the avoid condition and the subject's performance on the avoid condition.

Final discussion and conclusion

The task for ironic effects of attention suppression has been fully developed from the initial design, pilot testing, and complete analysis of the results. In this way, this task had different versions that used other stimuli (green square as Go stimulus, red square as NoGo stimulus), more time to respond to the stimulus (what made proportion of errors very small), and even a different cognitive task to evaluate the subjects performance, instead of a Go/NoGo task this was a stop-signal task which is similar but more challenging.

Our task to evaluate ironic effects of mental control is innovative and, to the author's knowledge, the first to evaluate ironic effects on the suppression of attention. This task can reveal important conclusions to whether we should actively avoid distractors when we are trying to concentrate or simply perform our task without worrying about the distractors.

Herein, ironic effects are assessed indirectly from the subjects' performance and not based on participants' self-report of the occurrence of ironic intrusions, as was the case for the initial studies on ironic effects of thought suppression[5,6]. This is a better method to evaluate the presence of ironic effects since it is more precise and less biased by the subjectivity of participants' experience. Moreover, the task has been carefully designed to avoid possible confounders and every factor on the task has been counterbalanced across the participants.

Our results suggest the occurrence of ironic effects due to the active suppression of the avoid images. However these effects are not always evident and could only be clearly identified in the group tested in an environment with some distractions (noisy testing). From the literature on the topic, we already knew that ironic effects of mental control arise when the subjects are under mental load, nonetheless, the Go/NoGo task that the subjects have to perform should be a sufficient mental load to produce ironic effects. It seems that the Go/NoGo task alone was not enough and only with the introduction of other distractions in the environment have ironic effects occurred. Moreover, our results suggest that in quiet environments the suppression of the distractor images on the avoid condition actually helped the subjects to concentrate, as their performance was better on the avoid condition than on the allow condition. On the contrary, in noisy testing conditions the active suppression of the avoid images worsen the subjects' performance relatively to the allow condition, indicating the presence of ironic effects. Maybe the initial version with a stop-signal task instead of a Go/NoGo task would be a better option to evaluate ironic effects of attention suppression on this task, because the stop-signal task is more challenging and cognitively more demanding than the Go/NoGo task.

The final small questionnaire given to the subjects was very useful to understand the results from the task performance. It revealed that the participants did not entirely follow the instructions and were not concerned about specifically avoiding the distracting images. In fact, the subjects suppressed the background images both on the avoid and the allow conditions and the effect of suppressing attention was not exclusive for the avoid condition. Telling the story about the images that help performance and the images that distract was to induce the active suppression of the images only on the avoid condition, but apparently our story was not convincing enough to our participants. In future studies with this task we might need to think about a more convincing story, or maybe the story we have told to these participants will be effective in other populations, like children or patients with obsessive-compulsive disorder.

Ironic effects were not entirely clear on this study, but that does not mean that our task is not a good experiment to evaluate ironic effects. This study was conducted in a group of control subjects and several other experiments of ironic effects in controls subjects have not been successful (Janeck & Calamari 1999; Muris et al. 1992; Smári et al. 1994). The ironic effects of mental control are more obvious or more prevalent in psychiatric populations, such as obsessive-compulsive patients or patients with eat-related disorders (e.g. anorexia nervosa), and our task can show stronger ironic effects of attention suppression on those psychiatric populations.

The analysis of ironic rebound effects showed a highly significant p-value for the interaction between the factors rebound, block, and images assignment, however the plots for rebound effect divided by Go frequency and image assignment did not show consistent results for the different dependent variables. This task was not initially designed to assess ironic rebound effects and, therefore, is not a robust experiment to evaluate their occurrence. Nevertheless, our results suggested that ironic rebound effects had occurred and it might be an important factor mediating the subjects' performance. A novel version of the task should be designed in order to better address this question, on which every participant would perform the allow condition before and after an avoid condition.

As a take home message, the results from our study of ironic effects of attention suppression suggested that, when working in a noisy environment, trying hard to concentrate and suppress the surrounding distractors might be a worse strategy than simply concentrate on our work/task and not worrying about the rest. However, in quiet environments the active suppression of possible distractors can effectively help us concentrate on our task.

Meanwhile, have you thought about the pink panther?

FUTURE WORK

The main goal of the development of the ironic effects of attention suppression task is to test the occurrence of ironic effects in neuropsychiatric disorders, specifically in obsessive-compulsive patients. Future studies will be conducted to test this task on those subjects, as well as on patients with attention-deficit/hyperactivity disorder. Symptom severity and psychological measures of mental control will be used to validate our results, like the completion of the White Bear Suppression Inventory (WBSI), in order to correlate task performance with the scores of those measures.

To understand the neurobiology and neuroanatomy of the processes studied here, further pharmacology and imaging studies need to be conducted. The theory of ironic effects of mental control was described in a broad sense but only very recently studies have started to explore the neural circuitries involved (Mitchell et al. 2007). Investigation on the neuromodulators that mediate the occurrence of ironic effects would be of extreme importance, since it could foster the development of new drugs to diminish the occurrence of unwanted thoughts in psychiatric disorders such as obsessive-compulsive disorder, anxiety, depression, eating, and sleeping disorders.

2

ATTENTION SHIFTING AND UNCERTAINTY

INTRODUCTION

It is not the strongest of the species that survives, nor the most intelligent;
it is the one most adaptable to change.

Adapted from Leon C. Megginson

Attention shifting is the process by which we shift our attention from currently relevant information to other information that has become more relevant due to a change in the environment. Shifting our focus of attention is an important part of mental flexibility and I will start by explaining mental flexibility and the different ways that it has been studied.

MENTAL FLEXIBILITY

Mental flexibility is a key factor for survival, since it equips us with the capacity to adapt our behaviour to appropriately deal with different situations and problems and successfully pursue our personal goals in a constantly changing environment. It represents our mental ability to shift between tasks (multitasking) or simultaneously think about multiple ideas. Mental flexibility plays a crucial role in learning, allowing us to behave adaptively in accordance with changing rules, goals, and demands[44].

Broadly, mental flexibility can be defined as the ability to shift a course of action or thought or to update previously learned behavioural strategies in adjustment to changing context contingencies[45,46]. It sets the capacity to change a previous response to an alternative that better suits the requirements of the situation, to abandon old habits for new behaviours whenever this adaptation is required by new set demands, and to cope with adverse events or unexpected outcomes. If a subject can successfully perform this adjustment then he/she is considered cognitively flexible. This flexibility is used in almost all situations and problems we face in our daily living, giving us the ability to change our problem-solving strategies in face of different challenges.

Oppositely, mental rigidity is the result of a deficient cognitive flexibility that can be found in many psychiatric disorders (obsessive-compulsive disorder[47], attention-deficit/hyperactivity disorder, autism[48], schizophrenia and bipolar disorder[49], Tourette's syndrome[50], anorexia nervosa[34,51]) or as a consequence of traumatic brain events. Patients that present mental rigidity are able to learn the rules that govern a certain environmental set and guide their behaviour accordingly to those rules, but lack the capacity to adapt their behaviour when the context changes and the rules are modified[1].

Cognitive flexibility has been studied using a variety of testing paradigms. In such paradigms, the participants initially have to learn to respond correctly to the stimuli presented. Once the participants have learned the rule of the game, a change in the rule occurs and participants must change their behaviour in order to discover the new rule and update their responses. These shifts in rule can occur several times during the task and the participants' performance on shifting trials is compared with performance on trials where there was no change in the rule. Performance on shifting trials is usually worse than performance on trials without shift[52].

Probably the most used paradigm to evaluate mental flexibility and set shifting in humans is the Wisconsin card sorting task[53]. On this task, the participant is asked to match test cards to reference cards according to the colour, shape, or number of stimuli presented on the cards. The participant is just told to categorize the cards based on one of those dimensions and not how to sort them. After each match, feedback is provided to indicate whether the matching rule the subject followed was the correct rule of classification (e.g., classification rules can be to match cards with (1) stimulus of the same shape, (2) stimulus of the same colour or (3) the same number of stimulus – see figure 2.1). From the feedback received, the subject learns the correct rule of classification. After the subject performs a certain number of correct matches, the classification rule is changed without warning and the participant will notice the change once he/she receives a negative feedback, indicating that the rule he/she has been using to match the cards is no longer the correct one. The subject must shift to a new rule of classification until he/she finds the correct one. The ability to shift the rule of classification after a negative feedback indicates that the participant is mentally flexible and capable of adapting his/her responses in face of changes in context. Psychiatric patients characterised by mental rigidity typically commit perseverative errors on this task, because they continue to follow the same rule even after receiving a negative feedback.

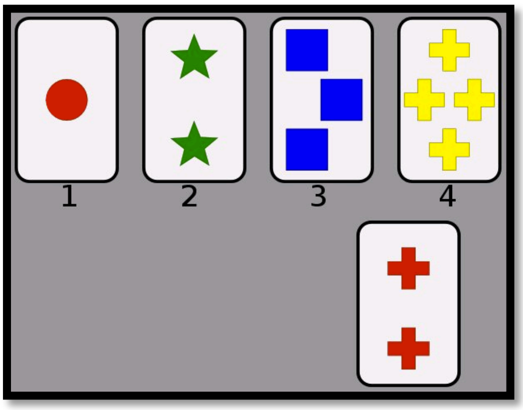


Figure 2.1 Example of a trial on a computational version of the Wisconsin card sorting task: the card to sort is the one that is not numbered (at the bottom) and it can be match with card 1,2, or 4, depending on the classification rule (same colour, same number of stimuli, or same symbol, respectively).
Source: http://en.wikipedia.org/wiki/Wisconsin_Card_Sorting_Test.

Besides its popularity, the Wisconsin card sorting task has some pitfalls, such as evaluating a variety of neurocognitive processes at the same time. Besides set-shifting, the Wisconsin card sorting task also evaluates processes like error-based learning, feedback processing and reinforcement learning, and working memory[54]. While testing different dimensions with one single cognitive task can be an advantage when evaluating psychiatric patients in a clinical context, it can be a drawback for scientific studies since it does not allow to disentangle the different cognitive processes involved on the task and to study them separately. Understanding which specific processes are affected in different psychiatric disorders is not possible using the Wisconsin card sorting task, since a poor performance on this task can be a result of impairments on different cognitive processes (e.g. reinforcement learning impairments can also result in poor task performance).

Mental flexibility is a vast area and involves several dimensions. Herein, we will study the ability to shift attentional focus, termed attention shifting.

ATTENTION SHIFTING

Directing our attention is essential to process efficiently whatever we are attending to. When talking with a friend in a noisy room we have to concentrate our auditory attention to our friend's words, otherwise we would be listening to all the noise in that room at the same time. Shifting attention implies changing our attentional focus to better process new relevant stimuli.

Usually, in attention-shifting tasks the predictive properties of the stimuli are deliberately changed without warning, in order to study how subjects shift and refocus their attention to new cues and adapt their behaviour. To succeed on this type of tasks, subjects have to set their attention to the relevant cue, but keep their behaviour flexible, so as to quickly adjust to new imperative or unexpected events. This ability is compromised in several neuropsychiatric disorders and it is important to establish a cognitive task to specifically test attention shifting without all the confounders present on the Wisconsin card sorting task.

Herein, we have developed an attention-shifting task and validated it in forty-five human adults. It is an adaptation of a task proposed by Yu and Dayan[55], which they have never tested empirically. This cognitive task is a combination of two paradigms: the Posner task and the linear maze navigation task.

THE POSNER TASK

The Posner paradigm was designed by Michael Posner[56] and has been one of the most influential paradigms in the recent decades in the study of attention.

In this paradigm, the participant has to detect and respond to a target that appears at one of two locations. The target is preceded by a cue that indicates its subsequent location. The task uses probabilistic cueing, where the cue explicitly predicts the location where the target will appear with a certain probability, termed cue validity. If the target appears in the direction indicated by the cue (valid cue trials) subjects will process the target more rapidly and accurately than in trials where the target appears in a direction which was not indicated by the cue (invalid cue trials). The difference in reaction times between invalid and valid trials (the validity effect, VE) increases with cue validity[57,58]. Usually, cue validity is set at 0.8, meaning that 80% of the trials will be valid and 20% will be invalid.

This design encourages the participants to set their attention towards the cue location, since that will be an accurate predictor of the subsequent target location for the majority of the trials and allows quicker target detection and responses (quicker reaction times) in valid trials.

THE LINEAR MAZE NAVIGATION TASK:

The linear maze navigation task is an example of an attention-shifting paradigm that has been developed to study perceptual attention shifting in rodents. In this task, rats learn to navigate in a maze following odour cues in order to reach a reward (food). After the animals have learned the odour cues, the cues change to visual ones and the rats can no longer rely on the odour to find their way to the food (the odour stimuli is still present but it is not predictive). The rats have to discover that a shift in the cues had occurred and learn to follow the visual cues that now guide them to reach the food.

Our task combines these two paradigms, as it uses probabilistic cueing to a subsequent target (as the Posner paradigm), but it also involves unexpected changes in the cue (like the linear maze navigation paradigm). Therefore, it was design to explore two forms of uncertainty: expected and unexpected uncertainty. Expected uncertainty describes the known unreliability of the cue, that is represented by cue invalidity (1 - cue validity; if cue validity is 80%, the cue invalidity will be 20%). Unexpected uncertainty represents unexpected changes in the environment that violate the subjects' predictions, that are represented by unexpected cue changes which will occur without notice along the task and cannot be predicted by the subjects[55].

NEUROBIOLOGY OF ATTENTION SHIFTING AND UNCERTAINTY

A considerable body of experimental evidence suggests that the cholinergic and noradrenergic systems relate to uncertainty, with acetylcholine being involved with expected uncertainty and norepinephrine with unexpected uncertainty.

On the Posner task, the validity effect (VE) has been observed to vary inversely with the levels of acetylcholine, which is consistent with the idea that acetylcholine reports expected uncertainty. If expected uncertainty is high, acetylcholine levels will increase and suppress the use of the cue.

In the linear maze navigation task, increasing norepinephrine levels in the rats with the drug idazoxan accelerates the detection of the cue-shift and learning of the new cues[59] and in equivalent experiences in monkeys and humans[60] cortical noradrenergic (but not cholinergic) lesions impair the shift of attention from one type of discriminative stimulus to another, which is consistent with norepinephrine being involved in reporting the unexpected uncertainty and mediating the ability to detect cue changes and shift to the new cue.

OBJECTIVE

This study aims to develop a novel cognitive task to evaluate attentional set-shifting in humans.

This new attention shifting task should specifically address the process of shifting attention without interference of other cognitive processes, as it happens with the existing tasks. Additionally, it is very important that this new developed task relies on well-established neuromodulators functioning, since the ultimate goal of its development is to study attention shifting in psychiatric populations and relate it with neurobiology of those disorders.

The objective of this part of my thesis was to fully develop this new task and verify if it is indeed a good measure of attention shifting.

METHODS

PARTICIPANTS

Forty-five adult subjects participated in this experiment. The majority of the participants were recruited from Universidade de Lisboa, either by direct approach or by email invitation.

All participants were native speakers of Portuguese and performed the task voluntarily after providing written informed consent.

PROCEDURE

After obtaining informed consent, the experiment was performed at the lab using a laptop on which the subjects completed the cognitive task. This task was programmed in Matlab using the Psychophysics Toolbox Version 3 (Psychtoolbox) to guarantee accurate timing for stimuli presentation and response collection.

EXPERIMENTAL TASK

The Attention shifting and Uncertainty task is an adaptation of a task proposed by Yu and Dayan in 2005[55]. Their task has never been tested empirically, so to the author's knowledge the present study is the first to empirically apply this cognitive task.

The attention shifting and uncertainty task was design to investigate the subjects' ability to set their attention towards a predictive cue and to update their attentional set whenever a cue shift occurs. Moreover, this task allows to explore two forms of uncertainty: expected uncertainty, the known unreliability of particular events within a familiar environment, and unexpected uncertainty, representing changes in the environment that could not be predicted and, thus, violate our top-down expectations[55].

On each trial of the task, two black arrows will appear on the centre of the screen: a horizontal arrow and a vertical arrow (see figure 2.3). Each arrow will be pointing in one of two directions: the vertical arrow can point up or down, while the horizontal arrow can point left or right. After the presentation of the arrows, a target appears. The target can appear in one of four locations (up, down, left or right).

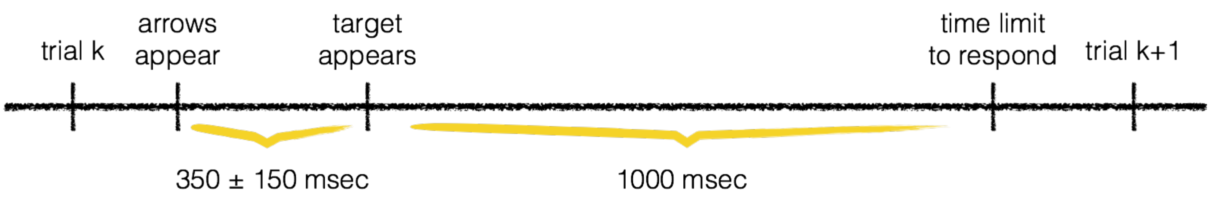


Figure 2.2 Timing of an illustrative trial on the task: when a new trial starts, the arrows are the first stimuli presented. After a variable delay of 350±150 milliseconds, the target is presented and the subject has 1000 milliseconds to respond to the target. Once the subject responds, a new trial will start, so the faster the subject's response, the faster the trial ends and a new one begins.

In the next 1000 milliseconds, the subject has to respond to the target location by pressing the corresponding arrow key on the keyboard (up arrow key, down arrow key, left arrow key or right arrow key, respectively). The main goal of this task is to indicate, as soon as possible, the correct location of the target.

One of the presented arrows (the cue or relevant arrow) will help the subject to predict the location where the target will appear by pointing to that location (Figure 2.3). The cue will help the subject to predict the target's location, but the cue is probabilistic and will only correctly predict the location of the target with a probability of γ ($\gamma >> 0.5$). γ is designated as the cue validity and, oppositely, $1-\gamma$ is the cue invalidity. This means that, e.g., if cue validity is 0.7, the cue will point to the direction where the target actually appears (valid trials) 70% of the trials, and the other 30% of the trials will be invalid (figure 2.3).

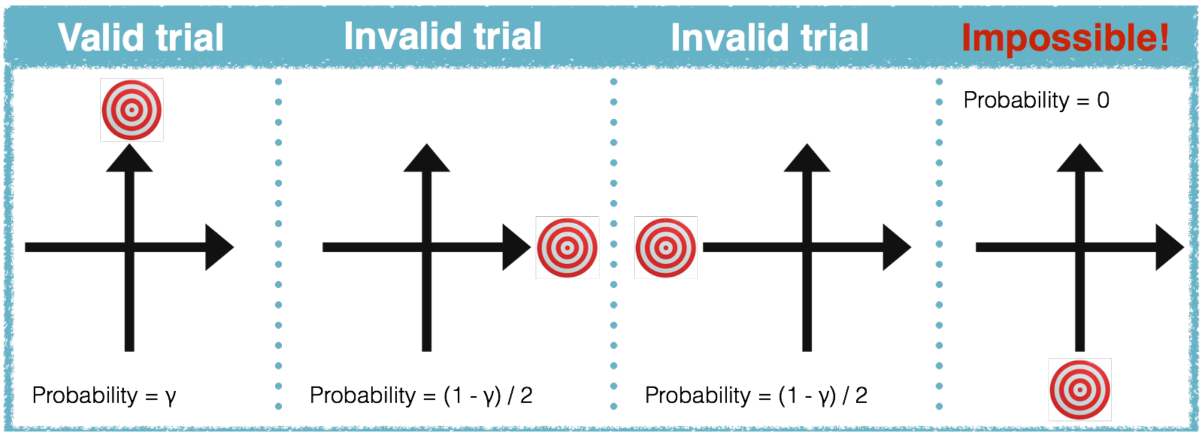


Figure 2.3 Different trial types on the task: considering the cue is the vertical arrow, from left to right, the first image depicts a valid trial, where the target appears on the location pointed by the cue; the second and third images represent invalid trials; and the fourth image represents a trial that can never occur, because the target can not appear on the axis of the cue (here the vertical axis) unless the trial is valid. This constriction was necessary to help the subjects to identify the cue. Probabilities of occurrence of each trial type are also indicated, where γ is the validity of the cue and, therefore, the proportion of valid trials and, inversely, $1-\gamma$ is the proportion of invalid trials.

On invalid trials, the target will appear in the direction of the other arrow (non-relevant arrow). For example, if the cue is the vertical arrow with $\gamma = 0.7$, 30% of the trials will be invalid and the target will appear on the horizontal axis (left or right). There are two ways that an invalid trial can occur (figure 2.3, images 2 and 3) and in one of those ways the non-relevant arrow is pointing to the target's location. To certify that the non-relevant arrow is not informative, it was constrained to point to the target's location only 50% of the times an invalid trial occurs. Overall, with a cue validity of 0.7, the target will appear 70% of the times on the location to where the relevant arrow is pointing (valid trial), 15% of the times to where the non-relevant cue is pointing (invalid trial – image 2, figure 2.3) and the rest 15% of the times to a location where none of the arrows is pointing (invalid trial – image 3, figure 2.3).

The subject has to discover which arrow is the cue to predict the location where the target will appear and then respond quicker to the target's location. However, throughout the experiment, the cue and its validity change with no warning to the participant. After a cue change, the trials continue similar to before but the predictive arrow now has changed. Since the cue is probabilistic, if it does not point to the target location on a trial does not mean that the cue has changed but probably that an invalid trial has occurred, therefore, the subjects cannot immediately notice that the cue has changed. Only after some trials will they realise that the cue they were following is not predicting the target location anymore. Along the task the cue will change four times and every time the cue changes a new block begins. Overall, the task is divided in five blocks that vary in size (blocks can have 30, 40 or 50 trials) and cue validity (0.7 or 0.8). The task design is described in table 2.1.

Table 2.1 Attention shifting and uncertainty task design.

Block	Number of trials	Cue validity
1	40 trials	80 %
2	30 trials	80%
3	50 trials	70%
4	50 trials	70 %
5	40 trials	80 %

Instructions

Subjects were instructed that one of the arrows is a cue to predict the target's location and that the cue can change without warning. They had to discover by themselves which arrow is the cue since the beginning of the task. Nonetheless, before starting the testing phase, subjects performed a short training, where any doubt regarding the task and the instructions was clarified.

DATA ANALYSIS

Data analysis was performed using R (R version 3.1.1), a software for statistical computation. The performance of the task was analysed in terms of reaction times and errors (omission and commission errors). However, due to the design of the task and the long time limit to respond (1 second), errors are very infrequent and less informative than reaction times. For that reason, the analysis of the task was extensively based on reaction times and errors were not so intensely exploited.

VALIDITY EFFECT

Reaction times were analysed by the calculus of the validity effect (VE). Validity effect is obtained as the difference between the mean reaction time for invalid trials and the mean reaction time for valid trials, as denoted in equation 6:

$$\mathrm{VE} = \mathrm{MRT}_{invalid} - \mathrm{MRT}_{valid} \tag{6}$$

Each block of the task was divided in groups of 10 trials and for each 10-trial group a validity effect value was calculated. Validity effects were then compared along the task and between the five blocks of the task.

Since this task is characterised by a continuous learning of the cue along each block, reaction times were also analysed continuously along the task. In order to do that, we choose to apply mixed-effects models to our data.

MIXED-EFFECTS MODELS

Mixed-effects models are statistical models that, like many other models, describe a relationship between a measure or response variable and the independent variables, which can be either experimental or observational variables that help to explain the response variable. Mixed-effects models differ from other models since they allow the use of both fixed and random effects in the same model. The fixed effects are terms of primary interest and are related to the experiment performed and its design; whistle random effects are drawn from variation that cannot be controlled by the design of the experiment. Wherein, the parameters of the task are modelled as fixed effects and the subjects' variation is modelled as random effects, permitting to take into account personal differences and general variability found among the participants of the study.

Therefore, mixed-effects models provide a flexible and powerful tool for the analysis of group data, especially for repeated measures, longitudinal, nested and unbalanced designs, since this tool is very robust against missing data, provided that data are missing at random.

The use of fixed and random effects can be seen as a hierarchically constructed model, considering one level for the subjects' effects (random effects) and another level for measurements within subjects (fixed effects). The statistical model is characterised in terms of two random variables: (1) the random variable $\mathcal{B}$, representing a q-dimensional vector of random effects, and (2) the random variable Υ, corresponding to an n-dimensional response vector. We can observe the value y of Υ but we cannot observe the value b of $\mathcal{B}$. In the model, we describe the unconditional distribution of $\mathcal{B}$ and the conditional distribution of $(\Upsilon|\mathcal{B})$ = b. The independent variables are used to estimate these parameters and to make inferences about them.

A linear mixed-effect model can be represented as:

$$\mathrm{y} = X\beta + Zb + \epsilon, \tag{7}$$

where y is the vector of observations, β is an unknown vector of fixed effects, b is an unknown vector of random effects, ϵ in an unknown vector of random errors, and X and Z are design matrices that relate the unknown vectors β and b to the vector of observations y.

The data from the task were fitted to four different models commonly used to explain behaviour (linear, logarithmic, exponential and power-law models) and random effects were considered for the intercept (estimating the variance representing how spread out the random intercepts are around the common intercept of each group) or for both random intercept and slope (additionally estimating the variance of random slopes around the common slope for that group).

We used four parameters in the models:

- trial, a continuous variable that represents the trial number within each block. The task has 190 trials that are divided in five blocks. Trial number within block is initialized every time a new block starts and ranges from 1 to 50, since 50 is the maximum number of trials that a block can have;
- type, a categorical variable with two levels that indicate whether that reaction time corresponds to an invalid (type = 0) or a valid (type = 1) trial;
- shift, which represents whether that block required a shift in the cue or not (only for the first block the shift is equal to zero, for all the other blocks shift is coded as one);
- validity, that represent the cue validity on that block and can either be 0.7 (70 %) or 0.8 (80 %).

Each model type (linear, logarithmic, exponential, and power law) was tested with different parameters. Starting from the simpler model to the most complex, we initially considered a model with the parameters type and trial and their interaction. Sequentially, the parameter shift was added, now having a model with parameters type, shift, trial, and all the interactions. A model with the parameters type, validity, trial, and interactions was also fitted. At last, the most complex model with all the parameters (type, shift, validity, trial) and their interactions was used. In the end, 32 models were used to fit the reaction times on this task.

Parameter estimation

The parameters of the models were estimated using the maximum likelihood technique.

If we think of a general model of ecological data as

$$y = f(x, P) + \varepsilon, \quad (8)$$

where y is a vector of values of a response variable, x is a vector of predictor variables, P is a vector of unknown parameters, and ε is a vector of errors. Function f can be any function we choose to fit the data (in our case, f will be a linear, logarithmical, exponential, or power-law function). We do not know the parameters P but we can guess the parameter values and calculate the distance from our predictions to the behavioural data. The residuals E are estimates of ε and are determined by the equation 9:

$$E = y - f(x, P). \quad (9)$$

The residuals can be used to calculate the parameter estimates. The maximum likelihood method often assumes that the residuals follow a normal distribution and the likelihood of any given residual is

$$L(E_i|P, \sigma) = \frac{e^{\left(-\frac{E_i^2}{2\sigma^2}\right)}}{\sqrt{2\pi\sigma^2}}, \quad (10)$$

where σ is the standard deviation of the residuals. Equation 10 gives the test for parameter estimates: if the resulting likelihood value is low, E is an unlikely estimate of ε and P is an unlikely estimator of the true parameters. The most likely values for the parameters are the ones that maximize the likelihood of E.

The total likelihood of all the residuals is calculated as the product of the likelihoods of the individual residuals:

$$\mathrm{L(E)} = \prod_{i=1}^{n} L(E_i|P, \sigma). \quad (11)$$

This product of the likelihoods is usually very large; therefore it is typical to work with the log of the likelihoods:

$$\log[L(E)] = \sum_{i=1}^{n} \log[L(E_i|P, \sigma)]. \quad (12)$$

To ease the computation we try to find the parameter estimates that minimize the negative log likelihood, that correspond to the same parameter values that maximize the likelihood and the log likelihood. Combining equation 12 with equation 10 and transforming log likelihood to a negative value, we have:

$$-\log[L(E|P,\sigma)] = \sum_{i=1}^{n} [\frac{E_i^2}{2\sigma^2} + 0.5\log(2\pi\sigma^2)]\,. \quad (13)$$

The maximum likelihood estimates are the parameter values that minimize equation 13.

Model comparison

The models created were compared in order to determine the model that best fits the behavioural data. This comparison was based on the model-fit statistics Akaike's Information Criterion (AIC)[61] and Schwarz's Bayesian Information Criterion (BIC)[62], derived from maximum likelihood parameters estimations.

Akaike's Information Criterion (AIC) is defined as

$$AIC\ (M) = -2logL(M) + 2 \cdot p(M), \quad (14)$$

where L(M) is the likelihood function of the parameters in the model M evaluated at the Maximum Likelihood Estimators and p is the number of parameters in the vector P of the model M. p reflects the complexity of the model and penalises models that have a higher number of parameters (the AIC imposes a penalty of two units per parameter in the model). The model that best fits the data is the one with the lowest AIC value. Deciding whether one model is significantly better than another model is somewhat arbitrary, but Burnham and Anderson[63] have suggested that a difference in AIC of 4 to 7 values corresponds roughly to 95% confidence that the lowest AIC represents a better model.

Another criterion to compare models is the Schwarz's Bayesian Information Criterion (BIC), calculated as

$$BIC\ (M) = -2logL(M) + (p(M) + 1) \cdot \log n, \quad (15)$$

where $L(M)$ is the likelihood function of the parameters in the model M evaluated at the Maximum Likelihood Estimators, p is the number of parameters in the vector P of the model M, reflecting the complexity of the model, and n is the number of data points. Also for this method, the model that best fists the data is the one with the lowest BIC value.

There has long been a fight between AIC and BIC criterions. Under some circumstances AIC tends to overestimate the model fit and can be a less secure measure than BIC. However, BIC over penalizes the complexity of the model and tends to choose the simpler models even when the more complex ones represent a better fit. To select the model that better represents the experimental data on this task, both AIC and BIC values will be taken into account.

RESULTS AND DISCUSSION

DEMOGRAPHIC DATA

Forty-five subjects completed the attention shifting and uncertainty task.

Participant sample characteristics:

All participants were adults, with ages ranging from 20 to 57 years old (mean age 29.29 ± 9.45). There were 30 females and 15 males (67% females).

PERFORMANCE ON THE ATTENTION SHIFTING AND UNCERTAINTY TASK

INCORRECT RESPONSES

The incorrect responses on the attention shifting and uncertainty task can be divided in two types: errors of omission, when the subject does not respond, and errors of commission, if the subject responded incorrectly to the target location. Proportions of each type of errors were calculated and represented in boxplots (figures 2.4 and 2.5).

Figure 2.4 Boxplot for proportions of omission and commission errors in the task. The data represent the distribution of the proportion of errors along the task made by all the subjects tested.

Omission errors are rare on this task (mean = 0.017 ± 0.064) because subjects were given 1 second to respond, which is enough time for them to press the response key even if they were distracted or confused by the occurrence of and invalid trial or a shift in the cue. On figure 2.5 (a), the proportion of omission errors is divided by trial type and even for invalid trials, omission errors are very small (practically zero).

Commission errors are more frequent than omissions. That is an expected result, since subjects can respond incorrectly to the target's location because they were focusing their attention on the location were the arrows were pointing. Nonetheless, commission errors are still insignificant (mean = 0.035 ± 0.051). Figure 2.5 (b) shows commission errors divided by trial type. There were more commission errors on invalid trials that on valid ones, which indicates a tendency to respond incorrectly when the target appears in a different location from the one pointed by the cue. To investigate whether subjects made a commission because they were following the cue, proportion of responses to the location indicated by the cue on invalid trials were calculated and are represented in figure 2.5 (b) as invalid trials: follow cue. In fact, some of the commission errors on invalid trials were responses to the location where the cue was pointing, but none of this is significant since proportion of commission errors is very small.

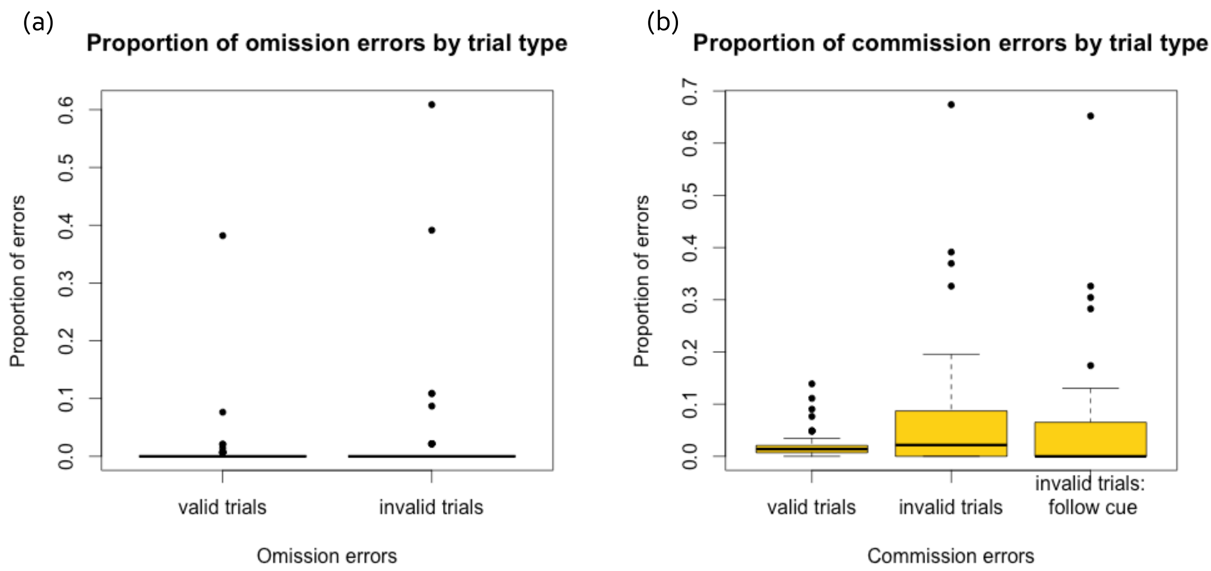


Figure 2.5 Boxplots for proportion of omission (a) and commission (b) errors divided by trial type. For commission errors a special case is introduced named invalid trials: follow cue; this represents a subset of commission errors made by the subjects on invalid trials for which they responded to the location where the cue was pointing.

VALIDITY EFFECTS

Validity effects (VEs) are calculated by the subtraction of the mean reaction time of valid trials from the mean reaction time of invalid trials for each sub-block (a sub-block is a group of 10 trials).

Hypotheses for validity effects

Hypothesis 1: The validity effect should always be positive or equal to zero:

- if $VE = 0 \Leftrightarrow MRT_{invalid} - MRT_{valid} = 0 \Leftrightarrow MRT_{invalid} = MRT_{valid}$, meaning that the subjects' reaction times are similar for valid and invalid trials. This should be found if the subjects are not following the cue to predict the target location and, therefore, whether the trial is valid or invalid is irrelevant for them.

- if $VE > 0 \Leftrightarrow MRT_{invalid} - MRT_{valid} > 0 \Leftrightarrow MRT_{invalid} > MRT_{valid}$, then the subjects are following the cue and correctly predicting where the target will appear, even before the target is presented on the screen. On the one hand, focusing the attention on the location to where the cue is pointing helps the subjects to process more rapidly the target once it appears (if it appears on that location – valid trial), and, therefore, respond faster. On the other hand, if the subjects are focusing their attention on the cue location and the target appears elsewhere (invalid trial), they will process the target more slowly since their expectations were violated. So this means that, when the subjects follow the cue, the validity effect should be positive because of two facts: (1) the reaction times for valid trials decrease, and (2) the reaction times for invalid trials increase.

- if $VE < 0 \Leftrightarrow MRT_{invalid} - MRT_{valid} < 0 \Leftrightarrow MRT_{invalid} < MRT_{valid}$. On a situation like this, the subject must simply be lost and following the non-relevant cue, not we do not expect this to happen.

Hypothesis 2: The validity effect should fall steeply when a shift in the cue occurs and increase along the block, until it reaches a ceiling. This means that the sub-block number 1 of blocks 2, 3, 4, and 5 should have a low validity effect (near zero), because the subjects are still following the previous cue/deciding whether the cue shifted of not. Once the subjects have noticed that the cue has shifted, they will start following the new cue and, along the block, rely more and more on the new cue, what will increase the value of the validity effect for the subsequent sub-blocks.

Hypothesis 3: Cue validity should affect the validity effect, since the higher the cue validity, the higher the confidence that the subject has on the cue. Blocks with cue validity 0.7 should have overall lower validity effects than blocks with cue validity 0.8.

Hypothesis 4: Block length can also affect the validity effect. On smaller blocks, the validity effects will not increase as much as in longer blocks, since the growth curve of validity effects values has been cut earlier.

Mean validity effects for the 45 subjects that completed the task were calculated and are plotted on figures 2.6, 2.7, and 2.8.

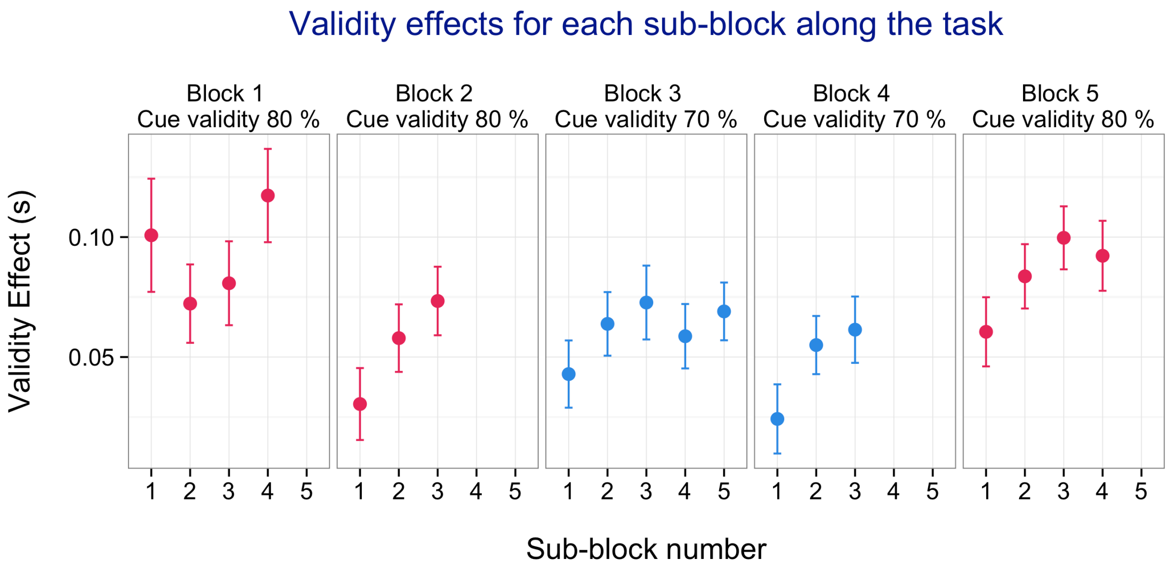


Figure 2.6 Mean validity effects for each sub-block (group of 10 trials) of the task. The colours indicate the cue validity, with blocks with cue validity 0.8 represented in red and blocks with cue validity 0.7 represented in blue. Error bars represent the standard error from the mean.

A two-way repeated measures ANOVA for the validity effects, with independent variables block, sub-block, and their interaction, showed a significant main effect of block (p-value = 0.0001) and sub-block (p-value = 0.0217).

The mean validity effects were always positive, meaning that, in general, the subjects did follow the cue to predict the target location. Moreover, after a cue shift, the validity effects were low and increase along the block (figure 2.8), and this is true for all the blocks on the task (figure 2.6).

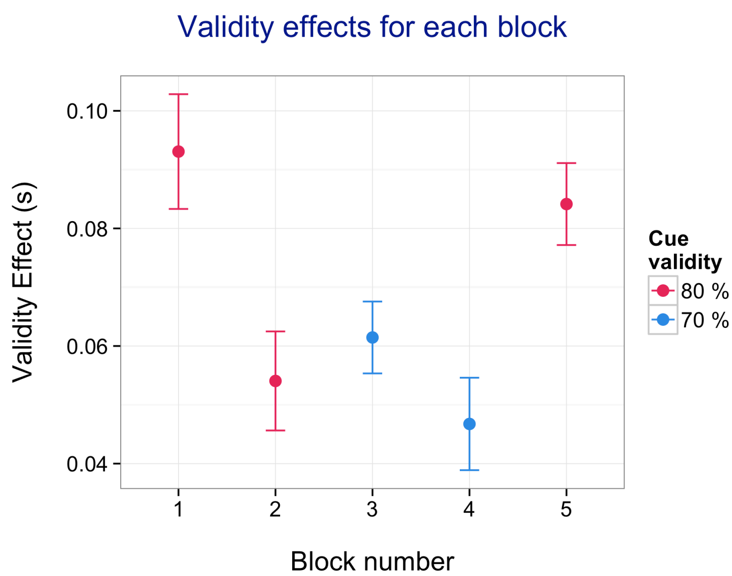


Figure 2.7 Mean validity effects for each block of the task. The colours indicate the cue validity, with blocks with cue validity 0.8 represented in red and blocks with cue validity 0.7 represented in blue. Error bars represent the standard error from the mean.

Blocks with lower cue validity (0.7) had lower validity effects than blocks with cue validity 0.8, and longer blocks had higher validity effects, respectively to the smaller blocks with the same cue validity, as expected. A two-way repeated measures ANOVA was performed to test the effect of cue validity and block length on validity effects; both the main effect of cue validity (p-value = 0.0125) and block length (p-value = 0.00015) were significant, as well as the cue validity x block length interaction (p-value = 0.0151). Looking at figure 2.6 and 2.7, the blocks number 1 and 2 do not seem to follow this tendency. In fact they are special cases in this task. Block number 1 is special because it is the first block of the task and does not start with a shift in the cue. This allows the subjects to rapidly learn which of the two arrows is the cue. Block 1 is the easiest block: it does not involve a shift and it has a high cue validity (0.8). The fact that the validity effects on this block are higher than from all the other blocks is then reasonable (notice that even the validity effect of the first sub-block on block 1 is higher than all the validity effects on the other blocks – figure 2.6). Block number 2 is also different as it is the first block where a shift in the cue occurs and the first time that subjects dealt with such a situation. That might be sufficient to make block number 2 harder. When the cue shifted for the first time it might have been harder for the majority of the subjects to understand what was happening (whether there were occurring a lot of invalid trials or the cue had actually changed) and that brought the validity effect of the first sub-block of block number 2 to a value near zero. From there on, the validity effect increased along the block but never reached the expected values for a block with high cue validity (0.8).

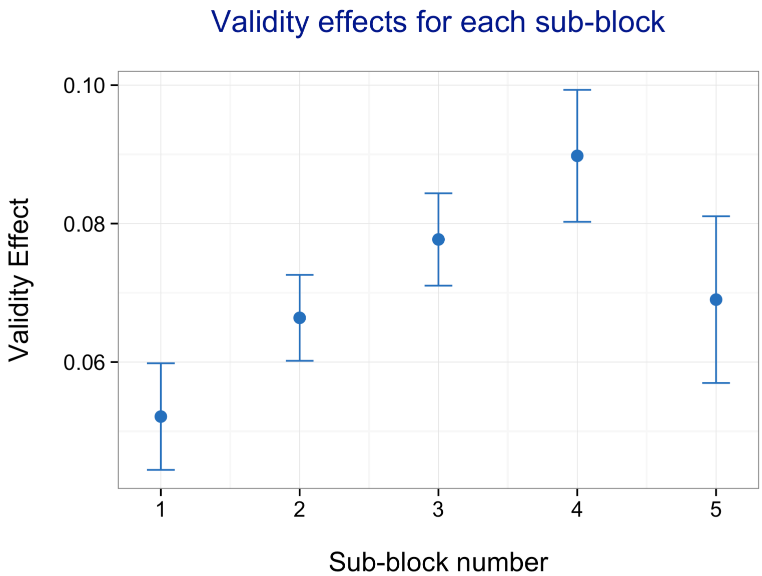


Figure 2.8 Mean validity effects for each sub-block of the task. This plot mimics a general block of the task and sub-block number indicates the progression along the block. Error bars represent the standard error from the mean.

Figure 2.8 shows the mean validity effect for each sub-block number and represents the evolution of validity effects along the sub-blocks of this task. A lower validity effect on sub-block 1 and subsequent increase along the block is evident, demonstrating that subjects rely more and more on the cue along the block. For sub-block 5 this is not true: sub-block 5 has a lower validity effect than sub-blocks 3 or 4. This is due to the fact that (1) on the entire task there is only one sub-block 5 (only the third block has five sub-blocks – 50 trials), so sub-block 5 has less data than the other sub-blocks and, consequently, higher variation, and (2) the single sub-block 5 of this task has cue validity 0.7, having a lower validity effect than the other sub-blocks that combine cue validities of 0.7 and 0.8 in their mean value.

MODEL FITTING ANALYSIS – MIXED-EFFECTS MODELS

Different models were fitted to the behavioural data, to better understand the dynamics of the reaction times of the subjects and the parameters that better explain them. To account for variability among the different subjects, mixed-effect models were applied, with fixed effects for the parameters of the model and random effects for personal differences among the subjects.

The fit of the models is described in table 2.2. For each model the function applied and the parameters used is indicated, as well as the defined random effects (random intercept and random slope). The value of log likelihood for all the models is given, which is a measure of the data likelihood for that model. The model that best fits the data is the one that maximizes the value of log likelihood. Additionally, both Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) values are provided. These values will be used to choose the best model, since they account for (1) the data likelihood and (2) the model complexity. The best model is the one that minimizes the value of AIC or BIC.

Hypotheses for model fitting

Hypothesis 1: Trial type (valid versus invalid) should be a highly significant parameter in any model and capture two distinct reaction times distributions:

- The values of the intercept for both trial type distributions should be similar, meaning that in the beginning of a block the subject will have similar reaction times in valid and invalid trials.
- The values of the slope should differ for the two trial types: valid trials will have a negative slope, since subjects will decrease their reaction times for valid trials along the block; while invalid trials should have a positive slope.

Hypothesis 2: Shift (no shift versus cue shift) should also be a significant parameter in any model, since there should be different dynamics in reaction times for the first block of the task, where no shift occurs, and the other blocks that are defined by a shift in the cue:

- For the first block (no shift), the intercept of both valid and invalid trials should be more separated with valid trials having a lower intercept value than invalid trials. This prediction is based on the fact that subjects rapidly capture which arrow is the cue on the first block and is also based on the high validity effect found for the first sub-block of the first block. Moreover, the slopes of the valid and invalid trials should have lower values in magnitude in the first block than for the blocks where the cue has changed.

- For blocks with a shift in the cue, the intercepts for valid and invalid trials should have very similar values, since this blocks start with a shift in the cue and on the first trials the subjects will still be following the previous cue.

Hypothesis 3: Cue validity can also be an important parameter to explain the data, affecting the slope of the two trial types:

- The higher cue validity (0.8) should predict higher values in magnitude for the slopes of the two distributions, relatively to the lower cue validity (0.7). This way, on blocks with cue validity 0.8, there should be a faster decrease on reactions times for the valid trials and a faster increase on reaction times for the invalid trials.

Hypothesis 4: Log likelihood values should be maximal for the more complex models. The likelihood of the model increases if the model can better predict the experimental data and the more parameters we introduce in the model the better the model predictions. Nonetheless, very complex models are not the best models to predict naturally occurring phenomena and every additional parameter we enter in a model should be strictly essential to explain the data. That is why log likelihood will not be used for model selection, but rather AIC and BIC that penalize the complexity of the model.

Table 2.2 Model comparison. 32 models were used to fit the behavioural data that differ in model type (linear, logarithmic, exponential, or power law), the combination of parameters used, and the defined random effects (only admitting random effects for intercept or additionally considering random effects for the slope).

Model	Factors	Random intercept	Random slope	BIC	AIC	log Likelihood
Linear	type, trial	Yes	No	-14955.0	-15011.0	7513.5
Linear	type, trial	Yes	Yes	-14946.2	-15051.2	7540.6
Linear	type, trial, shift	Yes	No	-14941.6	-15074.6	7556.3
Linear	type, trial, shift	Yes	Yes	-14858.3	-15173.3	7631.6
Linear	type, trial, validity	Yes	No	-14941.0	-15074.0	7556.0
Linear	type, trial, validity	Yes	Yes	-14807.6	-15122.6	7606.3
Linear	type, trial, shift, validity	Yes	No	-14749.1	-15092.1	7595.1
Linear	type, trial, shift, validity	Yes	Yes	-14026.4	-15069.4	7683.7
Logarithmic	type, trial	Yes	No	-14982.6	-15038.6	7527.3
Logarithmic	type, trial	Yes	Yes	-14960.3	-15065.3	7547.6
Logarithmic	type, trial, shift	Yes	No	-14971.1	-15104.1	7571.0
Logarithmic	type, trial, shift	Yes	Yes	-14856.7	-15171.7	7630.9
Logarithmic	type, trial, validity	Yes	No	-14959.8	-15092.8	7565.4

Logarithmic	type, trial, validity	Yes	Yes	-14800.1	-15115.1	7602.5
Logarithmic	type, trial, shift, validity	Yes	No	-14771.3	-15114.3	7606.1
Logarithmic	type, trial, shift, validity	Yes	Yes	-14017.7	-15060.7	7679.4
Exponential	type, trial	Yes	No	4557.3	4501.3	-2242.7
Exponential	type, trial	Yes	Yes	4590.6	4485.6	-2227.8
Exponential	type, trial, shift	Yes	No	4500.9	4367.9	-2165.0
Exponential	type, trial, shift	Yes	Yes	4617.1	4302.1	-2106.1
Exponential	type, trial, validity	Yes	No	4587.7	4454.7	-2208.3
Exponential	type, trial, validity	Yes	Yes	4729.4	4414.4	-2162.2
Exponential	type, trial, shift, validity	Yes	No	4665.7	4322.7	-2112.4
Exponential	type, trial, shift, validity	Yes	Yes	5369.8	4326.9	-2014.4
Power law	type, trial	Yes	No	4540.6	4484.6	-2234.3
Power law	type, trial	Yes	Yes	4559.0	4465.1	-2217.6
Power law	type, trial, shift	Yes	No	4483.7	4350.7	-2156.4
Power law	type, trial, shift	Yes	Yes	4612.3	4297.3	-2103.7
Power law	type, trial, validity	Yes	No	4574.8	4441.8	-2201.9
Power law	type, trial, validity	Yes	Yes	4731.9	4416.9	-2163.5
Power law	type, trial, shift, validity	Yes	No	4652.8	4309.8	-2105.9
Power law	type, trial, shift, validity	Yes	Yes	5373.5	4330.5	-2016.3

The exponential and the power law models fitted the data poorly. This might be an effect of the linearization of those models in order to fit them as linear mixed-effect models. The linearized exponential and power law models have an asymptote to zero and that forced the model to predict lower values for the reaction times that are far from the experimental data. Plots of the exponential model fit and the power law model fit can be found in appendix A.2.1. From now on, we will not discuss results relative to those two models, since they yielded much worse results than the linear and logarithmic models.

As predicted, log likelihood values are higher for the models with higher complexity. The best log likelihoods were found for the linear and the logarithmic models with parameters type, trial, shift, and validity and random effects for both the intercept and the slope. Nonetheless, considering also the model complexity with the AIC and BIC criterions for model comparison, the models with only type and trial, or additionally with the parameter shift, represented the best fits.

AIC and BIC were not congruent in the choice of the best model. While the best AIC values were found for the linear and the logarithmic model with parameters type, trial, and shift (highlighted values on the AIC column of table 2.2), with both random intercept and slope, BIC preferred the simpler logarithmic model (table 2.2, 9^{th} row) and, in second place, the logarithmic model with parameters type, trial, and shift, and only random intercept (table 2.2, 11^{th} row).

It is important to remember that AIC can overestimate the goodness of model fit, but in turn the BIC tends to over penalise the model complexity. Therefore, there is no rule to choose the better model in this case and some sensibility is required. First, we need to notice that the two best AIC values have a difference of -15173.3 - (-15171.7) = -1.6, which is not significant and indicates that the two models were equally good at predicting the behavioural data. Secondly, BIC has favoured the logarithmic models over the linear ones. In fact, the logarithmic models captured better the similar value for the intercept of the valid and invalid trials that can be seen in the behavioural data, as we were expecting (see figure 2.10). For those reasons, the logarithmic model was considered to be a better model to fit our data, relatively to the linear model.

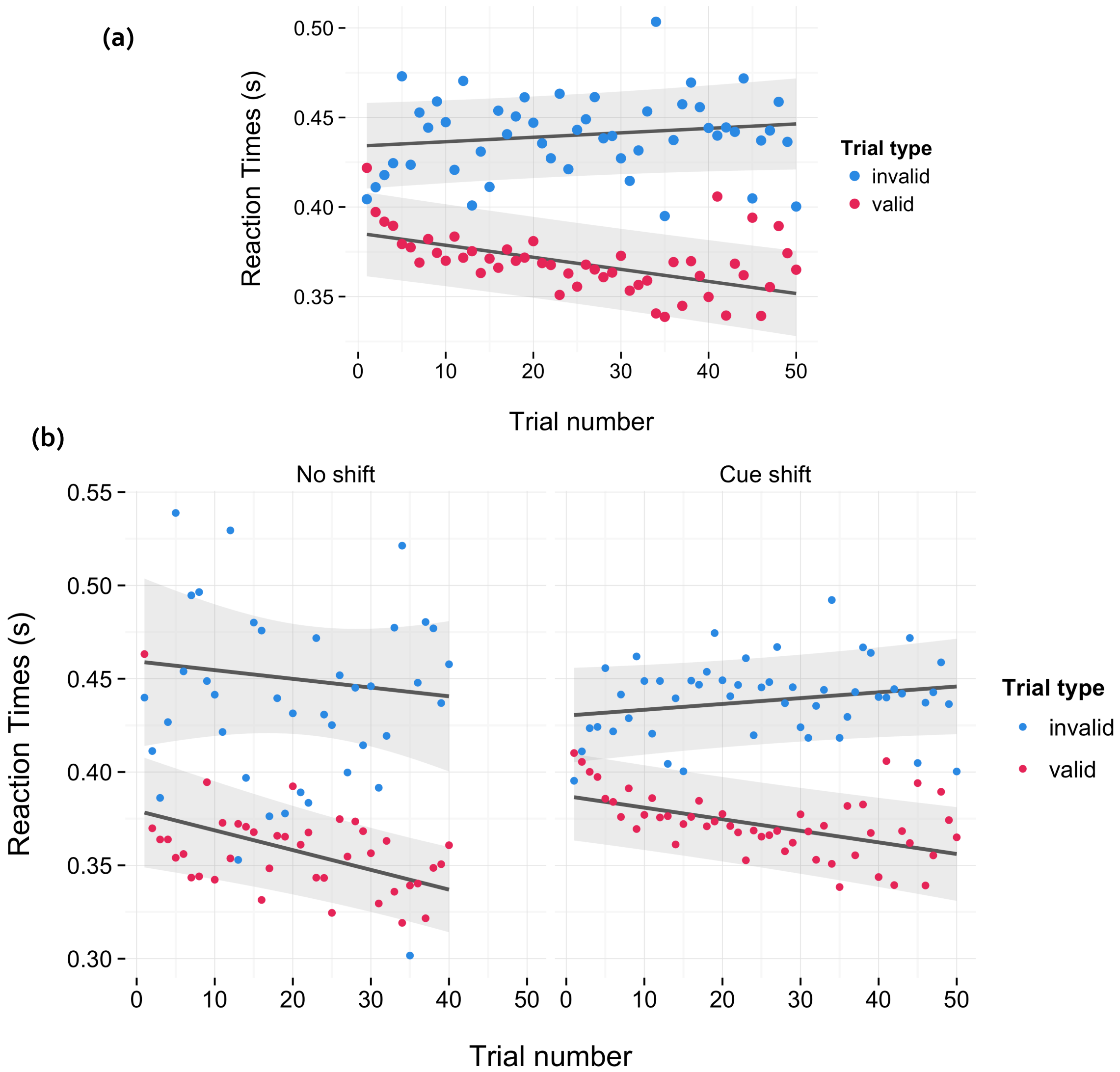
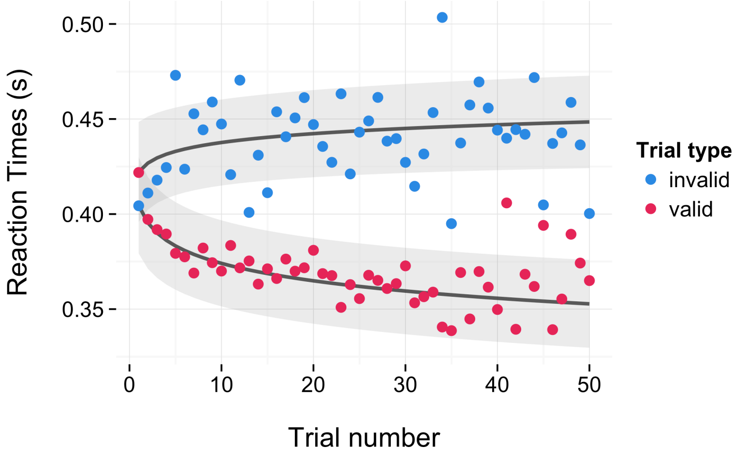


Figure 2.9 Linear model fits for the models with parameters (a) type and trial and (b) type, trial, and shift. Both models have random intercept and slope.

It remains to decide whether to include the parameter shift in the model. BIC significantly preferred the model without the parameter shift, nonetheless AIC considered shift an essential parameter. Since BIC over penalises the complexity of the model we consider that it is over penalising the addition of the parameter shift. In fact, the factor shift is essential to understand this specific task, since we want to study what happens when there is a shift in the cue. Therefore, the logarithmic model with parameters type, trial and shift is considered the model that best describes the task performance.

(a)



(b)

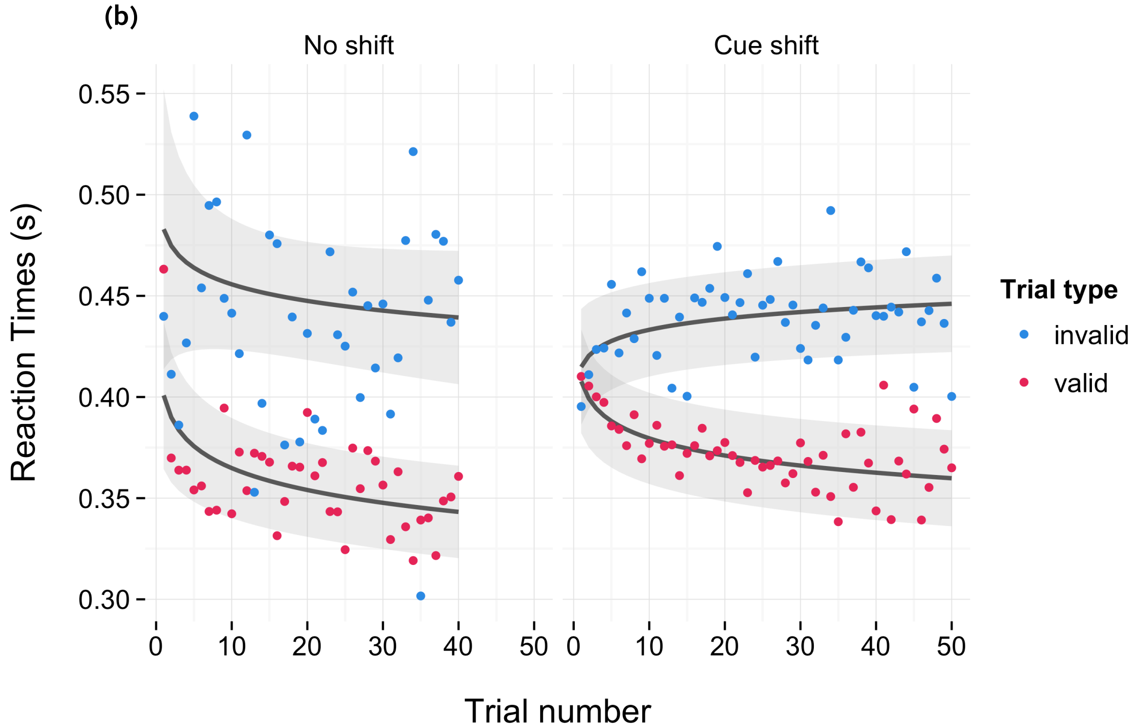


Figure 2.10 Logarithmic model fits for the models with parameters (a) type and trial and (b) type, trial, and shift. Both models have random intercept and slope.

Relatively to the random effects for the slope, there is no consensus to whether they should be included in the model. AIC had always preferred the models with random slope, while the BIC preferred the models with only random intercept. That is again due to the fact that BIC penalises model complexity more than AIC. We considered that random effects for the slope are very important, since subjects' variability affects not only the intercept, but also the slope of the reaction times. If a person has a higher capacity to shift the attention for the new cue, the slopes for the reaction times will be steeper. If the person takes several trials to understand that the cue has changed, the slopes will be less steep. In an extreme example, if the person has difficulty in shifting attention he/she can get lost during the task and stop following the cue, what would result in slope values near zero.

As expected, the reaction times follow distinct distributions for each trial type. Planned comparisons for the best model (the logarithmic model with parameters type, trial, and shift) were conducted with specific contrasts to assess the values for intercept and slope for each trial type: the intercept value for invalid trials is significantly higher than the intercept for the valid trials, both for the no shift block (intercept invalid – intercept valid = 0.0926, p-value = 0.000) and the cue shift blocks (intercept invalid – intercept valid = 0.0672, p-value = 0.000), the slopes for the valid trials are negative for the no shift block (-0.0136, p-value = 0.000) and the cue shift blocks (-0.0109, p-value = 0.000), indicating that the subjects reduced their reaction times along the blocks, while the slopes for invalid trials were non significantly different than zero for the no shift block (-0.0103, p-value = 0.788) and positive for the cue shift blocks (0,0071, p-value = 0.050).

Although the models with the parameter validity did not present the best fit, cue validity is an important factor to explain the data of this task. The fit of the most complex logarithmic model, with parameters type, trial, shift, and validity showed a significant main effect of type (p-value = 0.0000), trial (p-value = 0.0000), shift (p-value = 0.0023), and validity (p-value = 0.0001), as well as a significant effect for the two-way interactions type x shift (p-value = 0.0018), type x validity (p-value = 0.0223), and type x trial (p-value = 0.0000). Shift and cue validity are significant predictors of the reaction times and influence the performance in the task, as we have hypothesised.

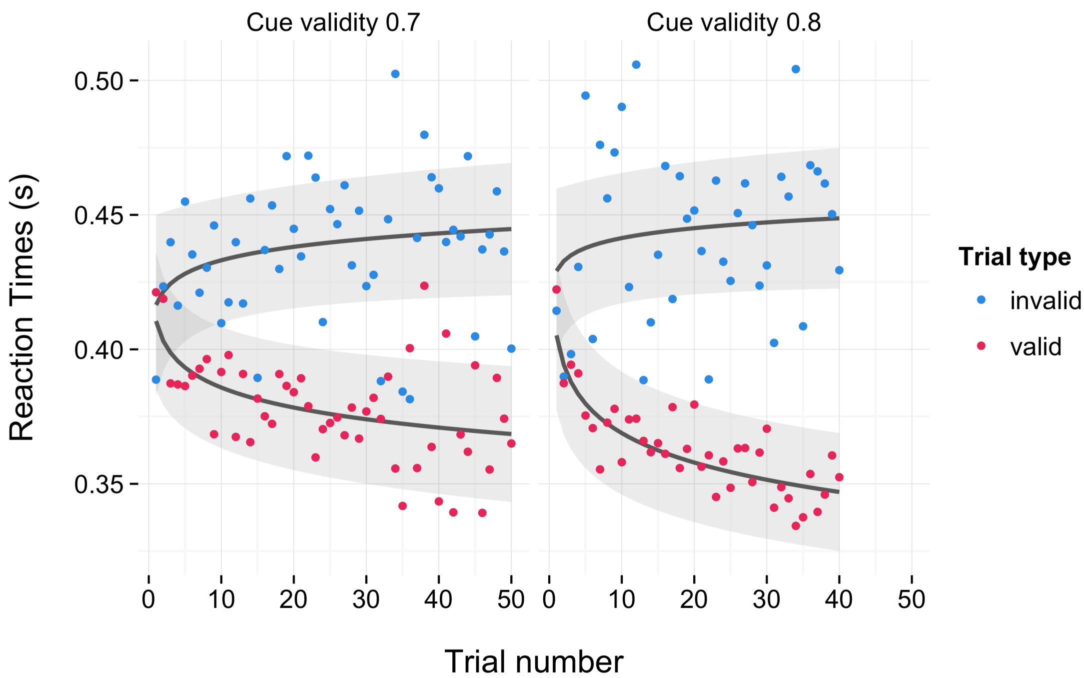


Figure 2.11 Logarithmic model fit for the model with parameters type, trial, and validity, with random intercept and slope.

Figure 2.11 shows the effect of validity on the reaction times. We can notice differences in the slopes for the valid trials for the two levels of validity, with a steeper slope for the higher cue validity (0.8). Nonetheless, the slope for invalid trials is not different for both cue validities. This is an effect of block 1, that has cue validity 0.8 and is influencing the intercept and slope values for the blocks with cue validity 0.8, forcing the intercept for valid and invalid trials to be separated and the slope for the invalid trials to be less positive (because block 1 has a negative slope for the invalid trials – figure 2.10).

ASSESSING THE FIT OF THE MODEL – RANDOM EFFECTS

The distribution of the 45 random effects corresponding to the 45 subjects that performed the task is depicted in figure 2.12. The value 0.0 represents the fixed effect for each parameter and the more spread are the random effects from zero, the more variability exists amongst the subjects. The random effects for trial (or interactions of trial with other parameter) represent the random effects for the slope and seem to be less variable than the random effects for the intercept.

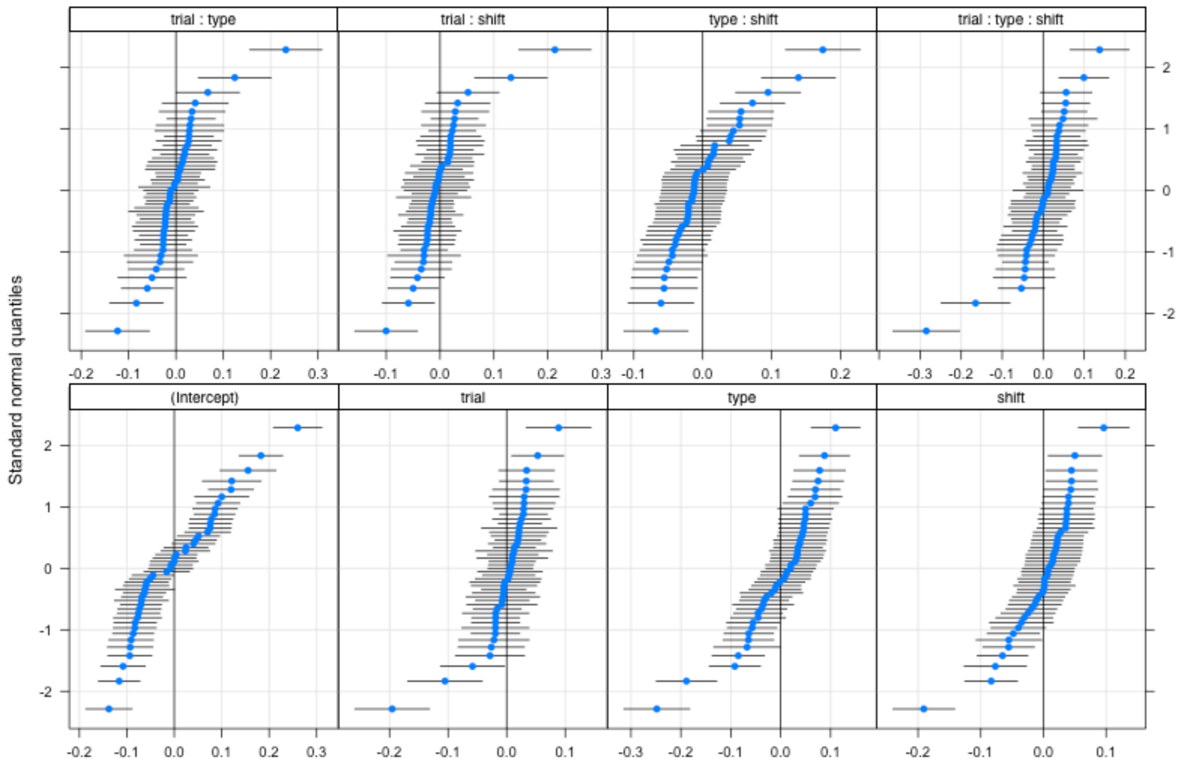


Figure 2.12 95% prediction intervals on the random effects versus quantiles of the normal distribution for the logarithmic model with parameters type, trial, and shift with random effects for the intercept and slope.

Final discussion and conclusion

A new cognitive task to evaluate attention shifting was successfully developed. The results from the analyses of task performance show that the participants were able to complete the task and to follow the cue, even after the occurrence of a shift. Nonetheless, not all subjects have a good performance. From the graphics shown in the results we get the idea that everyone can do this task successfully, but those graphs only represent the mean of the group of subjects. In reality, the majority of the participants was able to do the task and identify the changes in the cue, but some participants simply got lost and gave up on following the cue. Those participants only relied on the target to respond, they did not look at the arrows at all. Other participants reported that they would look at both arrows and prepare themselves to respond on any of the two locations to where the two arrows were pointing. More interesting is the fact that our analysis can perfectly capture those subjects that were not relying on the cue, both by the analysis of their validity effects or with the predictions from the mixed-effects models. Fitting mixed-effects models permitted to capture the differences in the subjects' performance and from the random effects in the model we can assess which subjects had better performance. In appendix A.2.2, subjects' differences on the reaction times along the block are shown. We can see that while some subjects have clearly distinct distributions of reaction times for the valid and invalid trials, others do not show this distinction and the intercept and slope for both trial types seem to be very similar.

This task can easily be completed without the aid of the arrows; to respond correctly it is enough to wait for the presentation of the target and then press the key that corresponds to the target's location. The cue is simply helping the subjects to respond quicker once the target appears. If the cue validity is small (near 0.5), the subjects feel that following the cue does not help them and stop relying on it. For that reason, the cue validity should never be lower than 0.7. Nonetheless, even using cue validities of 0.7 and 0.8, some subjects reported that they had not followed the cue. Probably those subjects felt more difficulty in following the cue after a shift and were not able to discriminate whether the cue has actually changed or if a lot of invalid trials were suddenly occurring. We cannot say with certainty that those subjects have an impaired attention shifting but we will conduct further studies to validate that hypothesis (studies where the subjects will perform the task and also complete a neuropsychological evaluation that assesses their mental flexibility).

Nonetheless, the group analysis is very encouraging. The group mean validity effects were always positive, meaning that in general the subjects did follow the cue to predict the target location. On previous versions of the task, VEs were frequently zero or even negative. That was taken as a signal that the task was too difficult and subjects got lost. Now, if the validity effects are positive for a group of 45 control subjects we can say that the task has an appropriate level of difficulty. It is important to remember that this task aims to test psychiatric populations, so the control subjects are expected to have a general good performance. Once we apply the task to patients with psychiatric disorders characterized by mental rigidity we expect to find lower validity effects, less differentiated reaction times for valid and invalid trials in the blocks where a cue shift occurred, and a higher proportion of errors.

The results from this study have corroborated our previous hypothesis. Nevertheless, some hypotheses were not met and those findings indicated important details on how the design of the task affects the subjects' performance. For example, we were not expecting to find a negative slope for the reaction times of the invalid trials on the first block (no shift block), however that finding suggests that the subjects can easily detect the cue on the first block, so easily that their first reaction times for the invalid trials are exaggeratedly high and decrease along the block because they get used to the unpredictability of the cue (cue invalidity).

We have developed a novel task to evaluate attentional shifting that has better characteristics than the already existing attention shifting tasks, considering that it does not involve other important cognitive processes, or at least it tries to minimize possible confounders. This task studies attention shifting under uncertainty and does not announce to the subject the change in context, but rather provides increasing evidence along the block that the cue has changed. Here, a good set-shifting capacity is crucial to detect the change and also to adapt the responses to the cue. On other existing tasks, the subject is explicitly told that the rule has changed[44,53] and the difficulty lies on the behavioural adaptation to the change. Actually, our task is much more similar to real life situations, because in most situations we are not told when a change in the environment has occurred and detecting those changes is difficult due to the existence of some expected uncertainty about the environment. Also, other tasks like the Wisconsin card sorting task, provide a feedback to the subject in the end of each trial. Impairments in feedback processing or reinforcement learning can result in perseverative errors, even if the subjects' mental flexibility is not affected. In our task, feedback is not necessary, since the subjects know whether they responded correctly or have made an error, and this excludes confounders of rewards or punishments processing during the performance. Taking again the example of the Wisconsin card sorting task, the participant founds that there was a change in the classification rule at the same time that it receives a negative feedback, so the unexpected event (shift in context) is always associated with a punishment and this relation can not be disentangled. When studying a psychiatric disorder or conducting imaging experiments simply using the Wisconsin card sorting task, one cannot say whether there is an impairment in reinforcement learning or in set-shifting.

The attention shifting and uncertainty task is also important to investigate noradrenergic functioning. Norepinephrine here works as an interrupting signal that codes unexpected changes in the context. Impairments in the noradrenergic system will negatively affect the performance of our task: for an under-functioning norepinephrine system, norepinephrine levels will be lower than normal and the subject will have an excess of confidence in the cue, resulting in increased difficulty in detecting cue shifts. Also the acetylcholine system plays an important role and can be investigated, since the levels of acetylcholine will code the cue invalidity; higher cue invalidity is represented by higher acetylcholine levels, what should lead the subject to rely less on the cue. An impairment on the circuitry of this neuromodulator will make the subject rely excessively on the cue and limit his/her flexibility to respond to invalid trials, which can result in reaction times higher than normal[55].

The attention shifting and uncertainty task here developed can also be related to the first part of this thesis about the ironic effects of mental control. For instance, the ability to shift our attention mediates our capacity to shift from one thought to another, what is an important ability to deal with stressors or cope with particularly difficult situations.

FUTURE WORK

In future studies to be developed at our lab, the attention shifting and uncertainty task will be applied to study impairments in attentional shifting in psychiatric populations, as attention-deficit/hyperactivity disorder and obsessive-compulsive disorder.

It is also important to validate this task with other measures of mental flexibility, a reliable psychological evaluation for attentional set-shifting, such as the Behavioural Rating Inventory of Executive Functions (BRIEF), and additional studies to investigate the effect of the neuromodulators norepinephrine and acetylcholine on the performance of the task should be conducted, using imaging or pharmacological methods.

Moreover, a computational model to analyse the performance of the task will be created. It will allow to analyse task performance using parameters that represent the functioning of the neuromodulators involved (norepinephrine and acetylcholine). The estimates of parameters will be related to the functioning of those neuromodulators' circuitries and detect whether impaired performance of the task is explained by deficient noradrenergic signalling or excessively high acetylcholine levels.

REFERENCES

1. Bissonette, G. B., Powell, E. M. & Roesch, M. R. Neural structures underlying set-shifting: roles of medial prefrontal cortex and anterior cingulate cortex. *Behav. Brain Res.* **250,** 91–101 (2013).
2. Wegner, D. M. When the antidote is the poison: ironic mental control processes. *Am. Psychol. Soc.* **8,** 148–150 (1997).
3. Wegner, D. M. *White bears and other unwanted thoughts: Suppression, obsession, and the psychology of mental control.* (Penguin Press, 1989).
4. Williams, A. D. & Moulds, M. L. Investigation of the indulgence cycles hypothesis of suppression on experimentally induced visual intrusions in dysphoria. *Behav. Res. Ther.* **45,** 2780–8 (2007).
5. Wegner, D. M. Ironic processes of mental control. *Psychol. Rev.* **101,** 34–52 (1994).
6. Wegner, D. M., Erber, R. & Zanakos, S. Ironic processes in the mental control of mood and mood-related thought. *J. Pers. Soc. Psychol.* **65,** 1093–1104 (1993).
7. Wenzlaff, R. M. & Wegner, D. M. Thought suppression. *Annu Rev Psychol*. **51**, 59–91 (2000).
8. Zukier, H. & Hagen, J. W. The Development of Selective Attention under Distracting Conditions. *Child Dev.* **49,** 870–873 (1978).
9. Wegner, D. M., Broome, A. & Blumberg, S. J. Ironic effects of trying to relax under stress. *Behav. Res. Ther.* **35,** 11–21 (1997).
10. Ansfield, M. E., Wegner, D. M. & Bowser, R. Ironic effects of sleep urgency. *Behav. Res. Ther.* **34,** 523–531 (1996).
11. Macrae, C., Bodenhausen, G., Milne, A. & Ford, R. On the regulation of recollection: The intentional forgetting of stereotypical memories. *J. Pers. Soc. Psychol.* **72,** 709–719 (1997).
12. Wegner, D. M., Ansfield, M. & Pilloff, D. The putt and the pendulum: Ironic effects of the mental control of action. *Psychol. Sci.* **9**, 196–199 (1998).
13. Cioffi, D. & Holloway, J. Delayed costs of suppressed pain. *J. Pers. Soc. Psychol.* **64,** 274–82 (1993).
14. Salkovskis, P. M. & Campbell, P. Thought suppression induces intrusion in naturally occurring negative intrusive thoughts. *Behav. Res. Ther.* **32,** 1–8 (1994).
15. Lavy, E. H. & van den Hout, M. A. Thought Suppression Induces Intrusions. *Behav. Psychother.* **18,** 251–258 (1990).
16. Clark, D. M., Ball, S. & Pape, D. An experimental investigation of thought suppression. *Behav. Res. Ther.* **29,** 253–7 (1991).
17. Clark, D. M., Winton, E. & Thynn, L. A further experimental investigation of thought suppression. *Behav. Res. Ther.* **31,** 207–10 (1993).
18. Roemer, L. & Borkovec, T. D. Effects of suppressing thoughts about emotional material. *J. Abnorm. Psychol.* **103,** 467–74 (1994).
19. Janeck, A. S. & Calamari, J. E. Thought Suppression in Obsessive-Compulsive Disorder. *Cognit. Ther. Res.* **23,** 497–509 (1999).

20. Muris, P., Merckelbach, H., van den Hout, M. & de Jong, P. Suppression of emotional and neutral material. *Behav. Res. Ther.* **30,** 639–42 (1992).

21. Smári, J., Sigurjónsdóttir, H. & Saemundsdóttir, I. Thought suppression and obsession-compulsion. *Psychol. Rep.* **75,** 227–35 (1994).

22. Wegner, D. M., Schneider, D. J., Carter, S. R. 3rd & White, T. L. Paradoxical effects of thought suppression. *J. Pers. Soc. Psychol.* **53,** 5–13 (1987).

23. Abramowitz, J. S., Tolin, D. F. & Street, G. P. Paradoxical effects of thought suppression: A meta-analysis of controlled studies. *Clin. Psychol. Rev.* **21,** 683–703 (2001).

24. Macrae, C. N., Bodenhausen, G. V., Milne, A. B. & Jetten, J. Out of mind but back in sight: Stereotypes on the rebound. *J. Pers. Soc. Psychol.* **67,** 808–817 (1994).

25. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24,** 167–202 (2001).

26. Braver, T. S., Reynolds, J. R. & Donaldson, D. I. Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron* **39,** 713–26 (2003).

27. Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L. & Snyder, A. Anterior cingulate cortex and response conflict: effects of frequency, inhibition and errors. *Cereb. Cortex* **11,** 825–36 (2001).

28. Kerns, J. G. *et al.* Anterior cingulate conflict monitoring and adjustments in control. *Science* **303,** 1023–6 (2004).

29. Mitchell, J. P. *et al.* Separating sustained from transient aspects of cognitive control during thought suppression. *Psychol. Sci.* **18,** 292–7 (2007).

30. Purdon, C. Thought suppression and psychopathology. *Behav. Res. Ther.* **37,** 1029–1054 (1999).

31. Tolin, D. F., Abramowitz, J. S., Przeworski, A. & Foa, E. B. Thought suppression in obsessive-compulsive disorder. **40,** 1255–1274 (2002).

32. Logan, A. C. & Goetsch, V. L. Attention to external threat cues in anxiety states. *Clin. Psychol. Rev.* **13,** 541–559 (1993).

33. Gotlib, I. H. & McCann, C. D. Construct accessibility and depression: an examination of cognitive and affective factors. *J. Pers. Soc. Psychol.* **47,** 427–39 (1984).

34. Cooper, M. J. & Fairburn, C. G. Selective processing of eating, weight and shape related words in patients with eating disorders and dieters. *Br. J. Clin. Psychol.* **31 (Pt 3),** 363–5 (1992).

35. Wegner, D. M. & Zanakos, S. Chronic thought suppression. *J. Pers.* **62,** 616–40 (1994).

36. Pennebaker, J. W. *Opening up: the healing power of expressing emotion*. (Guilford Press, 1997).

37. Freud, S. *"Notes upon a case of obsessional neurosis." The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume X (1909): Two Case Histories ("Little Hans" and the "Rat Man")*. (1955).

38. Freud, A. *The ego and its mechanisms of defense*. (London: Hogarth, 1937).

39. James, W. *The principles of psychology*. (New York: Holt, 1890).

40. Rassin, E. The contribution of thought–action fusion and thought suppression in the development of obsession-like intrusions in normal participants. *Behav. Res. Ther.* **39,** 1023–1032 (2001).

41. Green, D. M. & Swets, J. A. *Signal detection theory and psychophysics*. (Wiley, 1966).

42. Davison, M. C. & Tustin, R. D. The relation between the generalized matching law and signal-detection theory. *J. Exp. Anal. Behav.* **29,** 331–6 (1978).

43. Brown, G. S. & White, K. G. The optimal correction for estimating extreme discriminability. *Behav. Res. Methods* **37,** 436–449 (2005).

44. Bolton, H. M., Burgess, P. W., Gilbert, S. J. & Serpell, L. Increased set shifting costs in fasted healthy volunteers. *PLoS One* **9,** e101946 (2014).

45. *Encyclopedia of Clinical Neuropsychology*. (Springer New York, 2011).

46. Butts, K. A., Floresco, S. B. & Phillips, A. G. Acute stress impairs set-shifting but not reversal learning. *Behav. Brain Res.* **252,** 222–229 (2013).

47. Fineberg, N. A., Blackwell, A. D., Ph, D., Robbins, T. W. & Sahakian, B. J. Motor Inhibition and Cognitive Flexibility in Obsessive- Compulsive Disorder and Trichotillomania. *Am J Psychiatry*. **163(7)**, 1282-4 (2006).

48. Yerys, B. E. *et al.* Neural Correlates of Set-Shifting in Children With Autism. *Autism Res.* (2015).

49. Bozikas, V. P. *et al.* Controlled shifting of attention in schizophrenia and bipolar disorder through a dichotic listening paradigm. *Compr. Psychiatry* **55,** 1212–9 (2014).

50. Eddy, C. M. & Cavanna, A. E. Set-Shifting Deficits: A Possible Neurocognitive Endophenotype for Tourette Syndrome Without ADHD. *J. Atten. Disord.* (2014).

51. Lindner, S. E., Fichter, M. M. & Quadflieg, N. Set-Shifting and its Relation to Clinical and Personality Variables in Full Recovery of Anorexia Nervosa. *Eur. Eat. Disord. Rev.* **22,** 252–259 (2014).

52. Monsell, S. Task switching. *Trends Cogn. Sci.* **7,** 134–140 (2003).

53. Milner, B. Effects of Different Brain Lesions on Card Sorting. *Arch. Neurol.* **9,** 90 (1963).

54. Patel, D. D. *et al.* The neuropsychology of the schizo-obsessive subtype of schizophrenia: a new analysis. *Psychol. Med.* **40,** 921–933 (2010).

55. Yu, A. J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46,** 681–92 (2005).

56. Posner, M. I. Orienting of attention. *Q. J. Exp. Psychol.* **32,** 3–25 (1980).

57. Bowman, E. M., Brown, V. J., Kertzman, C., Schwarz, U. & Robinson, D. L. Covert orienting of attention in macaques. I. Effects of behavioral context. *J. Neurophysiol.* **70,** 431–43 (1993).

58. Downing, C. J. Expectancy and visual-spatial attention: effects on perceptual quality. *J. Exp. Psychol. Hum. Percept. Perform.* **14,** 188–202 (1988).

59. Devauges, V. & Sara, S. J. Activation of the noradrenergic system facilitates an attentional shift in the rat. *Behav. Brain Res.* **39,** 19–28 (1990).

60. Birrell, J. M. & Brown, V. J. Medial frontal cortex mediates perceptual attentional set shifting in the rat. *J. Neurosci.* **20,** 4320–4 (2000).

61. Sakamoto, Y., Ishiguro, M. & Kitagawa, G. *Akaike information criterion statistics*. (1986).

62. Schwarz, G. *Estimating the Dimension of a Model*. *Ann. Stat.* **6,** 461–464 (1978).

63. Burnham, K. P. & Anderson, D. R. *Model selection and inference*. S*pringer-Verla*g (1998).

Appendices

A. 1.1

Group 1 of images assignment for the task of ironic effects of attention suppression

A. 1.2

Group 2 of images assignment for the task of ironic effects of attention suppression

A. 1.3

Questionnaire about the performance on the task of ironic effects of attention suppression

Inquérito sobre a tarefa cognitiva

Agora que realizaste a tarefa cognitiva, gostaríamos de te fazer algumas perguntas.

É importante que respondas com sinceridade a cada pergunta, usando a opção que te parece mais próxima do que sentiste para cada situação.

Neste questionário não existem respostas corretas nem erradas. Lembra-te que o questionário é anónimo e não será usado para te avaliar.

ID do participante:

Durante a tarefa, em que medida tentaste olhar ou não olhar para os dois tipos de imagens?

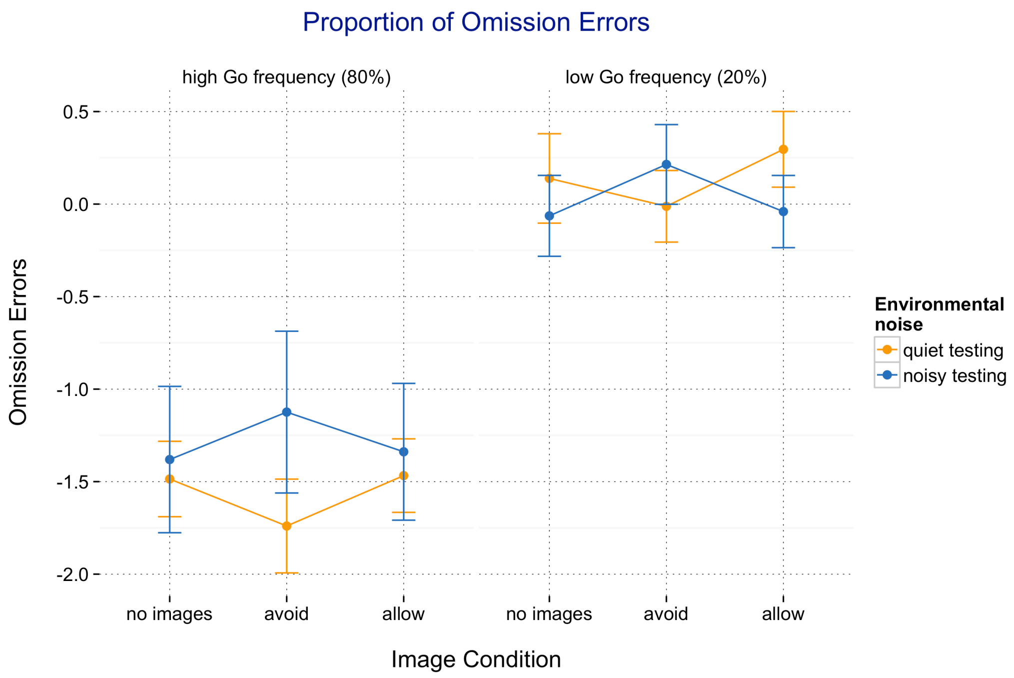
	Tentei muito não olhar	Tentei não olhar	Tentei um pouco não olhar	Não tentei olhar nem deixar de olhar	Tentei um pouco olhar	Tentei olhar	Tentei muito olhar
Imagens para ajudar							
Imagens para distrair							

Em que medida os dois tipos de imagens afetaram o teu desempenho na tarefa?

	Prejudicaram muito	Prejudicaram	Prejudicaram mas pouco	Não prejudicaram nem ajudaram	Ajudaram mas pouco	Ajudaram	Ajudaram muito
Imagens para ajudar							
Imagens para distrair							

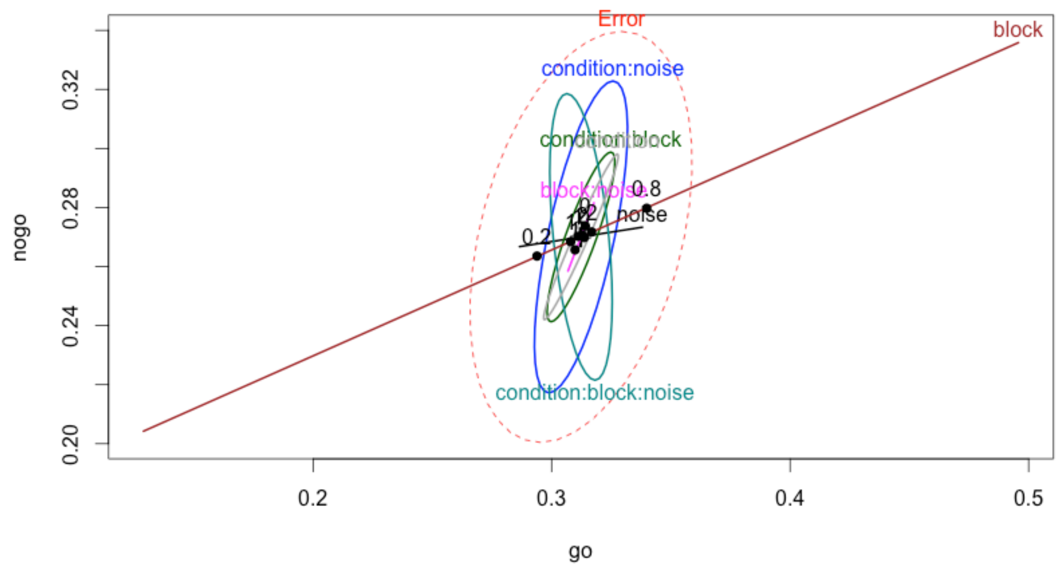
A. 1.4

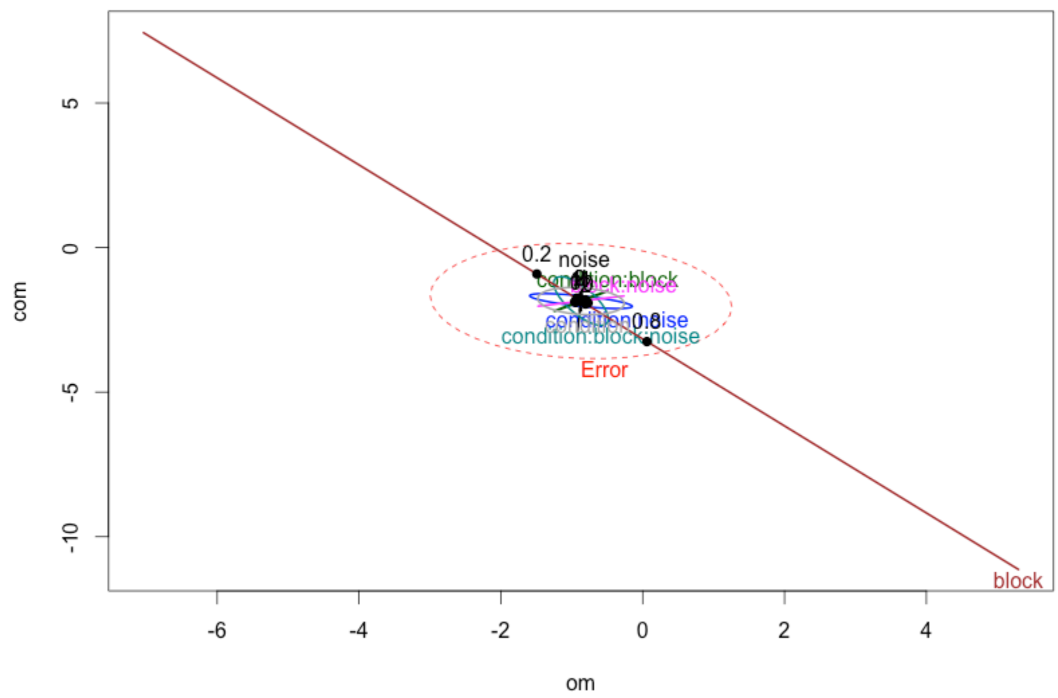
Mean values for the logit transformed omission and commission errors of the task of ironic effects of attention suppression. Mean values are divided by images condition, Go frequency and environmental noise; the error bars represent the standard error of the mean.



A. 1.5

Hypothesis-error plots for the ironic effects of attention suppression task. These plots show the hypothesis ellipses for the more important terms of the model for the response variables Go and NoGo mean reaction times (go and nogo, respectively) and omission and commission errors (om and com, respectively. Non-relevant terms were suppressed so that the plots could be readable.

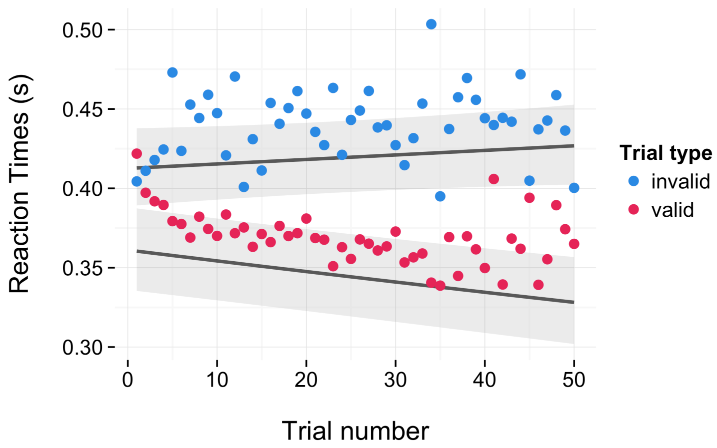




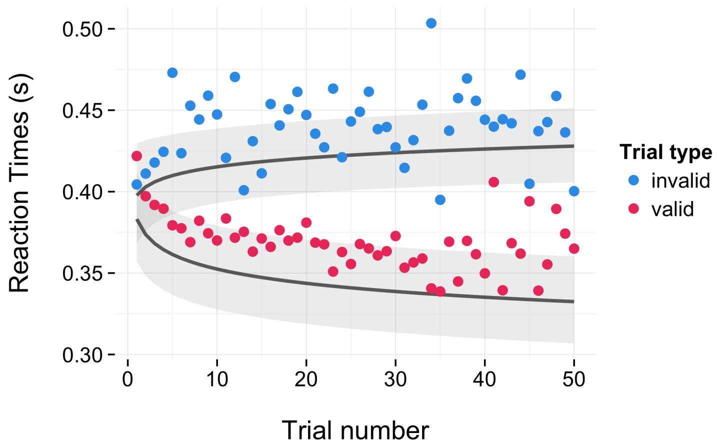
A. 2.1

Plots for the fit of the exponential (a) and the power law (b) mixed-effects models, with parameters trial type and trials and both random intercept and random slope, for the data of the attention shifting and uncertainty task.

(a)



(b)



A. 2.2

Predictions of the logarithmic model with parameters trial type and trial number, and both random intercept and random slope for each subject that completed the attention shifting and uncertainty task.